Proceedings of the 28th Annual Meeting of the International Society for Psychophysics

Ottawa, Ontario, Canada
October 18-21, 2012

Fechner Day 2012

Editors

Craig Leth-Steensen and Jordan Richard Schoenherr
Carleton University

Reference

Leth-Steensen, C., & Schoenherr, J. R. (Eds.) *Fechner Day 2012. Proceedings of the 28th Annual Meeting of the International Society for Psychophysics*. Ottawa, ON: The International Society of Psychophysics.

Acknowledgements

Fechner Day 2012 has been generously sponsored by:

Carleton University (Dean of the Faculty of Arts and Social Sciences and the Department of Psychology)

University of Ottawa (Dean of the Faculty of Social Sciences and the Department of Psychology)

Distribution

International Society for Psychophysics
c/o Craig Leth-Steensen
Department of Psychology, Carleton University
1125 Colonel By Drive, Ottawa, ON K1S5B6
Canada

Telephone: 613-520-2600 x 2254
Fax: 613-520-3667
Email: craig_leth_steensen@carleton.ca

Printed in Canada by:

Merriam Print (http://www.merriamprint.com/)

Cover design by Jordan Schoenherr

INTERNATIONAL SOCIETY FOR PSYCHOPHYSICS

Founded in Cassis, France, October 22, 1985

Executive Committee 2011-2012

President	Ana Garriga-Trillo, *UNED, Madrid, Spain*
Treasurer	William Petrusic, *Carleton University, Canada*
Secretary	Elisabet Borg, *Stockholms Universitet, Sweden*
Historians	Helen Ross, *University of Stirling, Scotland* Daniel Algom, *Tel-Aviv University, Israel*
Advisors to Committee	Mark Elliott, *National University of Ireland, Galway, Ireland* Lawrence Ward, *University of British Columbia, Canada*
Fechner Day 2012 Organizers	Steven R. Carroll, *St. Mary's University, Canada* Elena Gallitto, *Carleton University, Canada* Craig Leth-Steensen, *Carleton University, Canada* William M. Petrusic, *Carleton University, Canada* Jordan R. Schoenherr, *Carleton University, Canada* Dean A. Verger, *Carleton University, Canada*

PAST ANNUAL MEETINGS

YEAR	VENUE	ORGANIZER(S)
1985	Marseille, France	G. Canèvet & B. Sharf
1986	Cassis, France	A. M. Bonnet, G. Canèvet, C. A. Possamaï, B. Sharf, B. Berglund, & R. Teghtsoonian
1987	Durham, NH, USA	M. Teghtsoonian , & R. Teghtsoonian
1988	Stirling, UK	H. Ross, R. MacDonald, C. A. Possamaï, R. Teghtsoonian, M. Treisman, & R. Warren
1989	Cassis, France	A. M. Bonnet, G. Canèvet, C. A. Possmaï, & B. Sharf
1990	Würzburg, Germany	F. Müller
1991	Durham, NC, USA	G. R. Lockhead
1992	Stockholm, Sweden	G. Borg, Ǻ. Hellström, & G. Neely
1993	Mallorca, Spain	A. Garriga-Trillo, P. R. Miñón, C. García-Gallego, P. Lubin, J. M. Merino, M. J. Rubio-Gómez, & A. Villarino
1994	Vancouver, Canada	L. M. Ward
1995	Cassis, France	C. A. Possamaï, H. Ross, B. Scharf, R. Teghtsoonian, & L. M. Ward
1996	Padua, Italy	S. C. Masin
1997	Poznan, Poland	A. Gotebiewsky, E. Hojan, T. Hornowsky, P. Kokowsky, A. Majchrzak, P. Miecznik, M. Labowski, P. Pekala, A. Preis, E. Skrodzka, E. Wichlinska, & A. Wicher
1998	Québec, Canada	S. Grondin, Y. Lacouture, & R. Rousseau
1999	Tempe, AZ, USA	P. R. Killeen, & W. R. Uttal
2000	Strasbourg, France	C. Bonnet
2001	Leipzig, Germany	E. Sommerfield, T. Lachmann, & R. Kompass
2002	Rio de Janeiro, Brazil	J.A. da Silva, E. H. Matsushima, & N. P. Ribero-Philo
2003	Larnaca, Cyprus	B. Berglund, & E. Borg

2004	Coimbra, Portugal	A. M. Oliveira, M. Texeira, G. Borgess, & M. J. Ferro
2005	Traverse City, MI, USA	J. S. Monahan, S. M. Sheffert, & J. T. Townsend
2006	St. Albans, UK	D. E. Kornbort, R. M. Msefti, & A. W. MacRae
2007	Tokyo, Japan	S. Mori, T. Miyaoka, & W. Wong
2008	Toronto, Canada	B. A. Schneider, B. M. Ben-David, S. Parker, & W. Wong
2009	Galway, Ireland	M. A. Elliot, S. Antonijević, P. Berthand, P. Mulcahy, C. Martyn, B. Bargery, & H. Schmidt
2010	Padua, Italy	A. Bastianelli & G. Vidotto
2011	Raanana, Israel	D. Algom, D. Zakay, E. Chajut, S. Shaki, Y. Mama, & V. Shakuf

Preface:
Who is Fechner and Why He Still Matters

The publication of *Elemente der Psychophysik* by Gustav Theodor Fechner (1860) not only established the foundations of psychophysics as an area of inquiry but also set the stage for much of experimental psychology. While Fechner is one figure in the lineage of experimental psychology, his contributions set him apart from those that came before and after. Namely, although Fechner was influenced by the work of Ernst Heinrich Weber, the scope of Fechner's psychological research included not only the quantification of physical sensation and perceptual discrimination, but also natural history and consciousness (e.g., Fechner, 1851), evolution (Fechner, 1873), and the experimental study of aesthetics (Fechner, 1876). Yet the contributions of Fechner are not widely known within psychology as a whole (Scheerer, 1987) and we are confronted with the continuing challenge of demonstrating the connections between classic techniques and emerging areas of research.

As Boring (1950) notes, the impact of the publication of *Elemente* often overshadows Fechner's earlier works. Fechner first obtained a medical degree and, starting in 1821, published satirical evaluations of the medical science in his day. With his meagre income supplemented by translating over a dozen chemistry and physics texts, Fechner began to focus his studies on mathematics and physics. At the end of this decade, after becoming interested in the properties of electrical current, he published work on the measurement of direct current. Even at this early stage of his career, we see the physiological and the physical as recurrent themes in Fechner's work yet to be merged into what would later become psychophysics.

The crucial monistic synthesis occurred to Fechner while in bed on October 22, 1850. Mind and body were not dichotomous; they instead represented two aspects of reality. To explore this relationship, he spent the next decade developing the three methods of psychophysics (the methods of limits, constant stimuli, and adjustment) and what he referred to as *Weber's Law*. Though Weber's observations provided the groundwork for a ratio relating changes in intensity of an external stimulus to the changes detected by an individual, Weber himself identified neither the lawful relationship nor its complexity. It was Fechner's methods and theory that sparked investigations by his contemporaries and firmly established the psychophysical approach.

Having published *Elemente*, Fechner (1871) turned his attention to aesthetics, first examining the golden section and preference for rectangles and later culminating in the ill-fated attempt to authenticate two versions of Holbein's *Madonna with Burgomaster Meyer*. Experimental manipulations were clearly not possible, but Fechner reasoned that preference for each could be examined by soliciting the opinions of those who viewed the paintings. Although ultimately unsuccessful due to poor response rates and a biased sample, this attempt marked the beginnings of experimental aesthetics (Boring, 1950; Berlyne, 1971). Fechner (1876) again identified three new methods (the methods of choice, production, and use) and advocated a study of aesthetics "from below" rather than from philosophical principles (c.f., "from above") which culminated in the publication of *Vorschule der Ästhetik*.

Since Fechner's published *Elemente* and *Vorschule*, psychology has advanced into countless areas ranging from the study of complex social phenomena to the interaction of molecules in neuronal communication. Nonetheless, the relevance of psychophysics is still evident today. In this past decade alone, a search of the Web of Science identifies over 3,000 citations containing the term *psychophysics* in areas ranging from robotics and electronics to psychology and neuroscience. The search for a lawful relationship between the intensity of a stimulus and perception still guides the vast majority of research in neuroscience, perception, and cognition as well as many other related areas.

Psychophysical methods still provide the bedrock for experimental psychology with its origins as a "transdisciplinary research program" (p. 1211, Ehrenstein & Ehrenstein, 1999) still echoed in the comparatively recent efforts of cognitive science. In recognizing the significance and longevity of psychophysical theory within experimental psychology, we must also note the challenges presented by an approach developed over 150 years ago. Even though we might laud Fechner's integration of physiology and psychology, the objective and the subjective, we must also be cautious of adopting too narrow a focus. Fechner's approach "from below" represented a radical and necessary challenge to the approaches of his day that emphasized the opacity of personal experience. Nevertheless, we must also allow for the contributions of prior knowledge and top-down processes to the perceptual processes we observe in our controlled experimental settings. In this way, we will avoid relegating the contributions of Fechner and psychophysics to a historical footnote in psychology texts.

As the contributions to this year's Proceedings demonstrate, psychophysical theories and methods are still alive and well. Above all, these papers demonstrate that psychophysics and its history will have a bright future. We must, however, continue to demonstrate the relevance of these techniques to the next generation of psychologists so that they might extend that which Fechner has started.

In developing this conference, the organization committee for Fechner Day 2012 would like to thank the many contributors, those that have provided assistance throughout the planning of the conference, and the volunteers who will be helping to make the conference run smoothly. We are also extremely indebted to the financial support provided to this conference by John Osborne (Dean of the Faculty of Arts and Social Sciences at Carleton University), Ann Bowker (Chair of the Department of Psychology at Carleton University), Marcel Mérette (Dean of the Faculty of Social Sciences at the University of Ottawa), and Luc Pelletier (Chair of the Department of Psychology at the University of Ottawa).

Jordan Richard Schoenherr

References

Berlyne, D. E. (1971). *Aesthetics and Psychobiology.* New York: Appleton-Century-Crofts.

Boring, E. R. (1950). A History of Experimental Psychology. New York: Appleton-Century-Crofts.

Ehrenstein, W. H. & Ehrenstein, A. (1999). Psychophysical Methods, in U. Windhorst and H. Johansson (eds.) *Modern Techniques in Neuroscience Research.* Berlin: Springer, pp. 1211–1241.

Fechner, G. T. (1851). *Zendavesta, oder über die Dinge des Himmels und des lenseits*. Leipzig: Breitkopf und Härtel.

Fechner, G. T. (1860). *Elemente der Psychophysik*. Leipzig: Breitkopf und Härtel. Website: http://ia700404.us.archive.org/8/items/elementederpsych01fech/elementederpsych01fech.pdf.

Fechner, G. T. (1871). *Ueber die Aechtheitsfrage der Holbeirischen Madonna: Discussion und Acten.* Leipzig: Breitkopf & Härtel.

Fechner, G. T. (1873). *Einige Ideen zur Schöpfungsund Entwickelungsgeschichte der Organismen.* Leipzig: Breitkopf & Härtel.

Fechner, G. T. (1876). *Vorschule der Ästhetik*. Leipzig, Germany: Breitkopf & Härtel. Website: http://ia600305.us.archive.org/13/items/vorschulederaest12fechuoft/vorschulederaest12fechuoft.pdf.

Scheerer, E. (1987). The unknown Fechner. *Psychological Research*, 29, 197-202.

Table of Contents

In memoriam

Link, S., Narens, L., & Townsend, J. In memoriam: R Duncan Luce (1925-2012)…………. 1

Talks

Friday 9:00 – 10:45 AM session

Nizami, L. Loudness “fatigue” with two ears but not with one: Simultaneous dichotic loudness balance (SDLB) explained………………………………………………………………… 6

Nizami, L. The human cochlear mechanical nonlinearity inferred through the Schairer et al. (2003) model……………………………………………………………………………. 12

Adams, R., Lee, K-Y., Waddington, G., & Lee H-J. James McKeen Cattell and the method of constant stimuli in the psychophysics of movement……………………………………….... 18

Lee, K-Y., Adams, R., Lee H-J., & Waddington, G. Comparisons of indices of movement discrimination: Psychometric function, information theory, and signal detection analysis….. 24

Friday 11:00 AM – 12:45 PM session

Monahan, J., & Harke M. A. Applying SDT to a speeded paper and pencil test……………. 30

Link, S. Testing the logistic prediction of the relative judgment theory of psychophysical discrimination……………………………………………………………………………… 36

Cousineau, D. Does psychology require more (refined) data? ……………………………… 42

Friday 2:00 – 3:45 PM session

Zavagno, D., & Daneyko, O. The effect of non-adjacent luminance gradients on simultaneous lightness contrast…………………………………………………………………………… 48

Friday 4:00 – 5:00 PM session

Galanter, E., & Hannan, P. Consequential science: Psychophysical technologies and their implementations……………………………………………………………………….... 54

Saturday 9:00 – 11:15 AM session

Nakajima, Y., Ueda, K., Fujimaru, S., Motomura, H., & Ohsaka, Y. Acoustic correlate of phonological sonority in British English………………………………………………………… 56

Ueda, K., Noguchi, K., Satsukawa, Y., & Nakajima, Y. Perceptual roles of different frequency bands in Japanese syllable identification……………………………………………………….. 62

Kuroda, T., Grondin, S., Nakajima, Y., & Ueda, K. French and English rhythms are perceptually discriminable with only intensity changes in low-frequency regions of speech……………... 68

Avivi-Reich, M., Bae, M. H., Kang, Y. & Schneider, B. A. Listening to semantically anomalous sentences masked by noise and competing speech in a second language: A cross-language study on Korean-English bilinguals………………………………………………………………………. 74

Schoenherr, J. R., Logan, J., & Larose, C. Subjective confidence of acoustic and phonetic representations during speech perception……………………………………………………….. 79

Saturday 11:30 – 12:45 PM session

Bueno, J. L. O., & Nather, F. C. Implied movement perception in different static artworks affects subjective time…………………………………………………………………………………... 85

Townsend J. T., & Houpt, J. W. A new perspective on visual word processing efficiency…. 91

Saturday 3:30 – 5:00 PM session

Geissler H.-G., Vanagas, V., Svegzda, A., Bliumas, R., Stanikunas, R., & Vaitkevicius, H. Key-note lecture: Real-time gating of cyclic brain activity: Evidence from modulation of binocular rivalry induced by rapid stimulus interruptions…………………………………… 97

Sunday 9:00 – 11:15 AM session

Tristão, R. M., Garcia, N. V. M., de Jesus, J. A. L., & Tomaz, C. Psychophysical scaling in pain in newborns: Component comparison analysis…………………………………………… 113

Ross, H. E., & Schönbeck, G. L. J. Aristotle (c. 384-322 BC) and size-distance invariance.. 119

Wong, W. Audiovisual speech gaze strategies……………………………………………. 125

Hasuo, E., Nakajima, Y., Tomimatsu, E., Grondin, S., & Ueda, K. Perceiving filled vs. empty time intervals: A comparison of adjustment and magnitude estimation methods…………… 131

Leth-Steensen, C., & Chitty, A. Does requiring participants to produce the foreperiod duration eliminate the effect of foreperiod length on a subsequent RT task? ………………………... 137

Posters

Session I

P1. Borg E., & Borg G. Performance evaluation of diving using the BORG CR100 Scale®.. 143

P2. Parker, S. What's new? Not us………………………………………………………… 147

P3. Zellner, D. A., Hoer, K., & Feldman, J. Names will hurt you: Effect of label on liking and preference……………………………………………………………………………... 152

P4. Rooney, E., & Florentine, M. A. half-century's perspective on BÉKÉSY tracking and loudness growth at threshold……………………………………………………………… 156

P5. Masin, S. C. A brief trip into the history of psychophysical measurement…. 162

P6. Hellström, A. Cross-modality in comparisons of successive stimuli…………………….. 168

P7. Antonelli, M., & Zudini, V. From mathematics to psychophysics: David Hilbert and the "Fechner case"……………………………………………………………………………... 174

P8. Vicovaro, M.,& Burigana,L. The "launching effect" depends on size of colliding objects.180

P9. Bastianelli A., Vicentini M., Scandola M., & Fiorini P. Perception is nothing without control (of velocity)………………………………………………………………………... 186

P10. Scandola M., Bastianelli A., Vicentini M., & Fiorini P. The role of visual haptic delay in tele-operation protocols………………………………………………………………… 192

P11. Oliveira, A. M., Silva, A. D., Isidoro, M., Viegas, R. G., & Teixeira, N. S. Functional estimates of importance of facial features and oculomotor behavior: A straightforward or an intricate relation? ………………………………………………………………………….. 198

P12. Oliveira, A. M., Silva, A. D., Teixeira, N. S., & Viegas, R. G. What makes Mona Lisa smile? Investigating a halo model of configural face effects. 203

P13. Pereira, T., Oliveira, A., & Fonseca, I. B. The observable R and the unobservable r: Brain and rating responses in an integration task with pairs of emotional faces……………………. 209

P14. de la Sablonnière, N., Laflamme, V., Labonté, K., & Grondin, S. About the influence of perceiving disgusted, happy and sad faces on time perception……………………………… 214

P15. Laflamme, V., de la Sablonnière, N., Fournier, V., & Grondin, S. About the effect of fear and disgust on time perception………………………………………………………….. 220

P16. Schoenherr, J. R., Thomson, R., McGee, G., & Lacroix' G. A category-order effect? Subcategorical properties of stimuli determine a categorical effect……………………………… 226

P17. Schoenherr, J. R., & Lacroix, G. Dissociating categorization learning systems: Overconfidence and variations in learning rates in accuracy and confidence reports………... 232

Session II

P1. Wackermann, J., Pacer, J., & Wittmann, M. Temporal perception of acoustically presented time series…………………………………………………………………………….. 238

P2. Kattner, F., & Ellermeier, W. Can Pitch Be Measured On A Ratio Scale? ……………... 244

P3. Rammsayer, T. H., & Troche, S. J. Psychophysical approaches to neural efficiency and psychometric intelligence: Contrasting the coding efficiency hypothesis and the temporal resolution power hypothesis……………………………………………………………….. 250

P4. Verner M., & Rammsayer, T. How stable is the stimulus magnitude effect on perceived duration? ……………………………………………………………………………….. 256

P5. Hasuo, E., Kuroda, T., & Grondin, S. Assimilation between two neighbouring time intervals marked with tactile stimuli…………………………………………………………….... 262

P6. Miyaoka, T. Measurements of velvet hand illusion by magnitude estimation and paired comparison………………………………………………………………………………… 268

P7.Kuroda,T.& Grondin,S.Effect of space between fingers on tactile duration discrimination.274

P8. Toyoda, W., Doi, K., & Fujimoto, H. Influences of the sizes of tactile bars and dots on discriminability in people with visual impairment…………………………………………… 280

P9. Nather, F. C., & Bueno, J. L. O. The presentation of long-term duration of body movement in Impressionist artworks differently distort the perception of time. 286

P10. Nather, F. C., Fernandes, P. A. M., & Bueno, J. L. O. Timing perception is affected by Cubist paintings representing human figures. 292

P11. Nather, F. C., Mecca, F. F., & Bueno, J. L. O. Motion illusions caused by paintings of Op Art distort the perception of time. 298

P12. Schoenherr, J. R., Logan, J., & Larose, C. The identification of non-native speech sounds with psychophysical training……………………………………………………………… 304

P13. Runge, M., & Leth-Steensen, C. Effortful processing and the speed-accuracy tradeoff phenomenon………………………………………………………………………………….. 310

P14. Shakuf, V., & Mama, Y. Older and disillusioned: age differences in perceiving illusory line motion……………………………………………………………………………………... 316

P15. Shaki, S., & Fischer, M. H. Your neighbours define your value: Spatial bias in number comparison………………………………………………………………………………… 321

P16. Luccio, R., Fumarola, A., Tamburini, G., & Agostini, T. The spatial representation of non-symbolic numerical quantities……………………………………………………………….. 327

P17. Verger,D.A., & Petrusic,W.M. Extending SNARC: From single numbers to sequences. 331

R. Duncan Luce
(1925 – 2012)
Photo (2004) courtesy of Carolyn Scheer Luce

IN MEMORIAM

R. Duncan Luce
(1925-2012)

R. Duncan Luce, age 87, awarded the National Medal of Science in 2003, passed away in his sleep at his home in Irvine, California on August 11, 2012. He is survived by his wife Carolyn, daughter Aurora, her husband, and two granddaughters. At a family memorial he was eulogized by his friends and colleagues as a generous, thoughtful, scientist with a wry sense of very good humor and a happy family life.

Duncan Luce is best known in psychology for his formal approach to psychological theorizing. This development seems an unlikely outcome for a young student that entered the Massachusetts Institute of Technology at age 17, graduating three years later with his Bachelor of Science in Aeronautical Engineering. Following a short stint in the US Navy he re-entered MIT in mathematics and produced, in 1950, his PhD thesis entitled *On semigroups*. What seems so remarkable in retrospect is that his first publication in 1949, with A. D. Perry, was “A method of matrix analysis of group structure,” which appeared in *Psychometrika*. In 1950 he published “Connectivity and generalized cliques in sociometric group structure” also in *Psychometrika*. Semigroups, group structure, sociometric group structure: Is this the beginning of a career?

His amazing sequence of contributions to a more formal approach to psychological theorizing followed. The contributions may be thought of in several distinct and yet mutually influential groups. During the 1950’s his co-development of books and articles about mathematical principles applied to psychology yielded a basis for later works that must be viewed as some of the major contributions to the field in the 20th century. At the same time his contributions to mathematics, economics and measurement took full form. His influential *Games and Decisions* with Raiffa in 1957 set the stage for his wide recognition in Econometrics as did his previous *Econometrika* articles “Semiorders and a theory of utility discrimination” (1956) and “A probabilistic theory of utility” (1958) that preceded by one year his famous “little red book” *Individual Choice Behavior* (1959). These major accomplishments are only a refection of the amazing theoretical work created during the 1950’s.

His voluminous work in psychophysics begins in the 1950’s in a joint publication with Ward Edwards, “The derivation of subjective scales from just noticeable differences” appearing in *Psychological Review* (1958). In 1959 “A probabilistic theory of utility and its relationship to Fechnerian scaling”, and “Response latencies and probabilities” appeared in books and “On the possible psychophysical laws” in *Psychological Review*.

One of the most fundamental of Luce’s contributions, and a basis for much of his later thinking, was formalized in *Individual Choice Behavior*. This is the *choice axiom* (henceforth, simply AXIOM). The AXIOM is beguiling in its simplicity. In fact, on first reading of its representation in ICB, a novice would often require a bit of reflection to discern that it was not simply a statement of conditional probability theory. Omitting details such as when a choice object in a set has 0 probability of being chosen, we follow Luce in expressing the AXIOM as

$$P(R|X) = P(R|S) \cdot P(S|X) \text{ where } R \subseteq S \subseteq X.$$

The full AXIOM, despite its innocent appearance, bears the strong consequence that there exists a ratio scaled set of *values* on the objects $x \in X$ such that $v(x)$ is the *preference of* x_i and the probability of choosing x from any particular subset S of X, including X itself is

$$P(x|S) = v(x_i)/(\sum v(x_j),$$

where j runs over the index associated with the members of S, including x_i itself.

This formulation is sometimes referred to as the *strong utility model,* where the v's are interpreted as utilities. Moreover again, with provisions for dealing with probability-of-choice of 1 or 0 cases, the AXIOM is also equivalent to the *constant ratio rule,* which is defined by the property that $P(x|S)/P(y|S) = P(x, y)/P(y, x)$, where say, $P(x, y)$ is the probability that x is chosen over y when only the two are presented as alternatives. This strong but elementary appearing rendition of the AXIOM, especially in its strong utility formulation, sometimes misleads the reader into falsely supposing that the AXIOM is only the rule of normalizing a set of numbers so that they add to 1.

It is true that the normalization of positive numbers so that they sum to 1 (which in psychological context, I'll refer to as the "ratio of strengths" principle) was ubiquitous in science and math for quite a spell, to say the least. However, there are two main contributions that render choice theory a highly strategic psychological movement of the twentieth century. The essence of the first branch begins with Roger Shepard's seminal doctoral thesis where what is now referred to as the "similarity choice model" is given birth. The second branch lies within the mathematical structure of ICB.

In some ways it is more straightforward to start with the second branch: The AXIOM is substantially deeper and more consequential than a ratio of strengths quantity by itself. As the above modest descriptions intimates, and the details of ICB demonstrates, it implies a profound type of independence ("invariance" might be a preferable term) among choice objects, even as the set of such objects expands or contracts. The redoubtable constraints imposed on choice and preference behavior by this axiom are responsible for a wide terrain of decision and to some extent foundational measurement research in economics, psychology (especially decision making) and a number of other fields.

When the "strength of preference" value is represented as a product of a similarity value and a response bias value, the "similarity choice model" is created and the model moves from being applicable "only" to valenced choice objects, to sizeable arenas in perception and cognition (e.g., detection, identification, categorization, etc.), as pre-figured by Roger Shepard. However, the latter uses have almost always given up AXIOM per se, although retaining the "ratio of strengths" formulation which provides the normalization. This and other facets of the similarity choice model, its properties, and relationships to other models of identification were explored in a paper by Landon and Townsend, "An experimental and theoretical investigation of the constant ratio rule and other models of visual letter recognition" *Journal of Mathematical Psychology* (1982). Rob Nosofsky and Doug Medin, to name two major theoretical figures in the field of categorization were instrumental in developing models of categorization and general classification based on the Luce-Shepard choice theory.

These days, mathematical modeling is increasingly carried out within rather complex, if cogent, quantitative theories using extraordinarily powerful computer software. This is all to the good and the field must and will evolve. Nonetheless, the kind of beautiful analytic modeling incorporating classical definitions-axioms-theorems-proofs-linkage-of-theorems-and-structure-to-experimentation, which can be seen and honored in so much of R. Duncan Luce's theorizing will remain part of scientific psychology and cognitive science.

The well-known summer institutes at Stanford University on mathematical psychology united a cadre of now famous psychological scientists including Richard Atkinson, Robert Bush, Bill Estes, Eugene Galanter, Duncan Luce, Patrick Suppes, and so many others. These institutes, started in the 1950s and carried into the 1960's, helped develop the many seminal publications on the applications of probability to psychological theorizing. Many edited books brought to an excited graduate student audience the most recent works in the field. In 1960 Luce edited "*Developments in Mathematical Psychology*", and with Bush and Galanter in 1963 and 1965 *Handbook of Mathematical Psychology* Vols. 1, 2 and 3 and the two volumes of *Readings in Mathematical Psychology* in 1963 and 1965.

The discussions that occurred during these summer meetings set the stage for the development of new ideas about measurement. Classical measurement theory dates back to Helmoltz (1887) and Hölder (1901). Its application to the measurement of physical quantities is based on the idea that there are an observable ordering relation and an observable binary operation on the objects to be measured that satisfy certain testable axioms. The most important of such axioms say that a binary operation is commutative and associative.

In the first half of the twentieth century these ideas of classical measurement theory dominated measurement theory. Measurement theorists considered as unsound the methods of psychological measurement that were beyond simple ordinal measurement, because such methods were not implemented through an observable commutative and associative operation. In response S. S. Stevens formulated his famous theory of measurement in *Science* in 1948. In the 1950's, he and his students systematically applied his theory, particularly to psychophysical phenomena. This was the state of the scientific measurement environment when Luce began his work on measurement.

Luce's 1956 article "Semiorders and a theory of utility discrimination" was a forerunner to a new theory of measurement called "the representational theory of measurement" that Luce and Patrick Suppes later developed. Luce and Tukey's seminal article "Simultaneous conjoint measurement: A new type of fundamental measurement" appeared as the first article in the new *Journal of Mathematical Psychology* and demonstrated that some psychological qualities can be measured in a manner just as rigorously as qualities in the physical sciences.

They considered the situation of an ordering relation $\geq$ on a Cartesian product $X \times Y$ so that X can be measured by an interval scale ψ_1 and Y can be measured by an interval scale ψ_2 such that for all (x, y) in $X \times Y$,

$$(x, y) \geq (u, v) \text{ iff } \psi_1(x) + \psi_2(y) \geq \psi_1(u) + \psi_2(v).$$

Luce and Tukey provided conditions about how X and Y observably interacted with each other in terms of $\geq$ so that the interaction induced a commutative and associative operation.

The idea of using observable structure to define a commutative and associative operation that may not be directly observable, and using classical measurement with that operation to measure observable qualities and reinterpret the results in terms of the observable structure, became the key concept in Volume I and part of Volume II of Krantz, Luce, Suppes, and Tversky's seminal work, *Foundations of Measurement*. Eventually three volumes totaling 1426 pages appeared. These are considered a milestone of twentieth century science.

Luce and Narens generalized the key ideas of Vols. I and II of *Foundations of Measurement* so that the induced operation could be non-commutative or non-associative. They called their theory "non-additive measurement theory." Because they had no scientific application of it, they published it in the *Journal of Pure and Applied Mathematics* in 1976. Over a ten year period, Luce and Narens and students worked on this project. Finally in 1985 all the important pieces of the theory were developed and Luce and Narens published a 72 page summary in the *Journal of Mathematical Psychology*. At the end of that article they included an application to utility theory that provided alternative psychological explanations to some key phenomena of Kahneman's and Tversky's prospect theory. The research on non-additive measurement theory is summarized in Volume III of *Foundations of Measurement.* Luce (2000) employed non-additive measurement theory as the basis for his theoretical and experimental masterpiece *Utility of Gains and Losses: Measurement-Theoretical and Experimental Approaches*.

Duncan's other books are marvels of scientific accomplishment. The early work with Raiffa, *Games and Decisions* (19570, the very famous *Individual Choice Behavior* (1959), the series of three *Foundations of Measurement* volumes with Suppes, Krantz and Tversky (1971-1990), *Response Times* (1986), *Sound and Hearing* (1993) and the just mentioned *Utility of Gains and Losses* are great works deserving the attention of students and scientists alike.

As a leader in theoretical developments Duncan also became a leader in the political development of scientific psychology. From 1964-1968 and 1971-74 he was a member of the Mathematical Social Science Board where he served as Chairman 1966-68 and 1972-74. He was elected to the National Academy of Sciences in 1972 and served in various capacities during the 1980's. He was a Fellow or Member of such prestigious scientific societies as the American Association for the Advancement of Science, the American Psychological Society, Federation of Behavioral, Psychological, and Cognitive Sciences where he served as Vice President, 1984-87 and President, 1988-91. He was also a member of the Psychometric Society serving as President, 1976-77, a member of the Psychonomic Society and its Representative to Federation of Behavioral, Psychological, and Cognitive Sciences, 1983-85. As a member of the Society for Judgment/Decision Making he served on the Executive Committee, 1987-90. One of the Societies he helped found and sponsored was the Society for Mathematical Psychology where he served on the Executive Committee, 1978-80, and as President, 1979.

His awards and honors speak to a career of accomplishment and distinction. He was elected to the Society of Experimental Psychologists in 1963, the American Academy of Arts and Science in 1966, he received the distinguished Scientific Contributions Award from the American Psychological Association in 1970. In 1972 he joined the National Academy of Sciences and in 1986 received the American Association for the Advancement of Science's Prize for Behavioral Science Research. In 1994 he was elected to the distinguished American Philosophical Society and in 2001 received the Gold Medal Award for Life Achievement in the Science of Psychology from the American Psychological Foundation. He was awarded the National Medal of Science in 2003 and President G. W. Bush presented the award to him in 2005. In 2004 Duncan received the Society of Experimental Psychologists Norman Anderson Award for Lifetime Contributions to Psychology, and in 2012 he received the Patrick Suppes Award in Psychology from the American Philosophical Society.

Here was a breathtaking career peopled by great scientists advancing ideas about psychology, economics, and scientific measurement itself. At the same time Duncan remained readily available to students and colleagues alike. Always an advocate for excellence, a strong voice for scientific psychology and a willing collaborator with a delightful wit, he was a model of decorum, a man of genius whose great works inspire others to continue his view of a scientific psychology.

Stephen Link
University of California, San Diego

Louis Narens
University of California, Irvine

James Townsend
Indiana University

LOUDNESS “FATIGUE” WITH TWO EARS BUT NOT WITH ONE: SIMULTANEOUS DICHOTIC LOUDNESS BALANCE (SDLB) EXPLAINED

Lance Nizami
Independent Research Scholar, Palo Alto, California, USA
nizamii2@att.net

Abstract

In SDLB (1950 onwards), the loudness contribution from the continually exposed “fatiguing” ear is matched by adjusting the intensity of an intermittent stimulus at the other (“comparison”) ear. The latter intensity declines, indicating “loudness fatigue”. However, the loudness of a continuous well-supra-threshold stimulus to one ear (with the other in quiet) does not diminish. Here is a quandary, presently resolved through a novel model dependent upon (1) the aforementioned non-fatiguing, and (2) the olivocochlear bundle, which “turns down the volume” in the ear opposite to one experiencing stimuli. The model explains how “fatigue” varies with stimulus variation, revealing “fatigue” as an SDLB artifact.

The papers cited here include crucial early contributions to SDLB. As such, some are quoted, to capture the atmosphere of the work while explaining it succinctly.

SDLB allegedly measures “the decrease in the loudness of a steady acoustic stimulus *during* its presentation” (Egan, 1955, p. 111; original italics), called perstimulatory fatigue. Note SDLB’s purported motivation (Small, 1963, p. 289):

> If a pure tone is presented to a listener continuously and at the end of five minutes he is asked if the stimulus sounds differently than it did in the beginning, his usual response is “no, it sounds the same”. The perceived loudness of the stimulus remains very nearly unchanged. It is as though the listener had neither an internal loudness standard nor an effective memory and thus is able to compare the loudness in a particular segment of time only with the loudness of the stimulus in the immediately preceding segment – an imperceptible change. The key to the perception and measurement of a loudness decrement under these circumstances seems to be the availability of a comparison stimulus.

SDLB uses a comparison stimulus. Egan (1955, p. 111) explained while introducing jargon:

> A fatiguing stimulus having constant spectral characteristics is presented to one ear. A comparison stimulus whose intensity the listener can control is simultaneously presented to the other ear. During the simultaneous dichotic stimulation the listener adjusts the intensity of the comparison stimulus until it appears as loud as the fixed, fatiguing stimulus. After this loudness balance the comparison stimulus is turned off, but the fatiguing stimulus continues to sound. Later the comparison stimulus is again briefly presented for a loudness balance with the fatiguing stimulus. In this way the temporal course of the decline in loudness of the fatiguing stimulus may be obtained.

Unfortunately, the meaning of “fatiguing” and “comparison” has sometimes been reversed. Further, “test ear” has been used for either ear. Here, in an attempt at clarity, the ear receiving the “fatiguing” stimulus will be called “ipsilateral” and the ear receiving the comparison will be called “contralateral”. The terms “fatiguing” and “comparison” will still be used when needed. A single experimental “run” in classic SDLB was described by Egan (1955, p. 112):

> The temporal sequence of the stimuli in measuring perstimulatory fatigue was as follows. The fatiguing and the comparison stimuli were presented together for 20 seconds, during which time the listener adjusted the intensity of the comparison stimulus for a loudness balance. Both stimuli were then turned off and the listener called out his [attenuator] setting. Forty seconds later both stimuli were presented again for another loudness balance. After cycle was repeated several times, the fatiguing stimulus was left on. During this fatiguing period, the comparison stimulus was presented every minute for 20 seconds beginning on the minute. The recovery from perstimulatory fatigue was traced by turning off both the fatiguing and comparison stimuli for 40 seconds and then presenting both stimuli for another loudness match.

The comparison periods respectively preceding and following “perstimulatory” were deemed “prestimulatory” and “poststimulatory”. Figure 1 (after Egan, 1955) shows these three stages. The per- or post-stimulatory “fatigue” indicated by a matching comparison stimulus intensity is the latter’s dB SPL subtracted from the average prestimulatory comparison dB SPL.

Small and Minifie (1961, p. 1028) noted that “Unfortunately, it takes an appreciable interval to obtain a loudness balance”; Egan’s (1955) subjects admittedly used all of each of their allotted 20 sec. Also, a subject’s attenuator was always set to its minimum between loudness matches, and further, an arbitrary amount of attenuation, unknown to the subject, was introduced by the experimenter. Thus, “As a consequence of the attenuation introduced into the [attenuator] pads of the experimenter and observer, on any given [loudness] balance the intensity [*sic*] of the comparison stimulus at its onset was either completely inaudible or relatively weak” (Thwing, 1955). Each subject was thus obliged to begin each adjustment session by raising the intensity of the comparison stimulus.

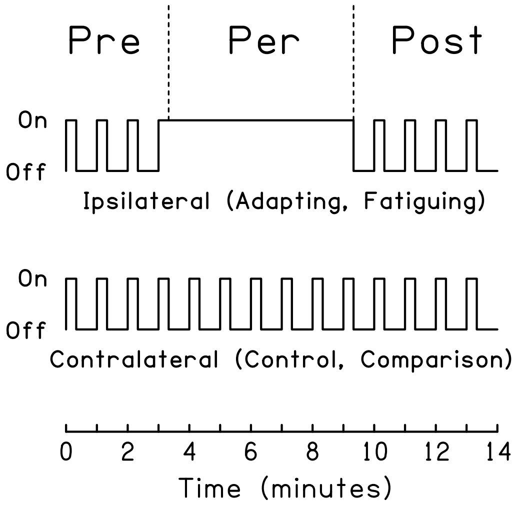


Figure 1. Stimulus schedule during a typical SDLB run (see text).

“Fatigue”: two ears versus one

In SDLB, perstimulatory ipsilateral “fatigue” increases with ipsilateral stimulus duration, although its rate-of-change decreases with time such that it appears to asymptote (e.g., Hood, 1950; Carterette; 1955; Egan, 1955; Thwing, 1955; Jerger, 1957; Small & Minifie, 1961; Sergeant & Harris, 1963; Fraser, Petty, & Elliott, 1970; Petty, Fraser, & Elliott, 1970; Stokinger, Cooper, & Meissner, 1972). Time-to-asymptote appears to be at least 5 min, and perhaps more than 10 min, for stimuli ≥80 dB SPL, and it increases with ipsilateral intensity (Hood, 1950; Egan, 1955; Carterette, 1955; Jerger, 1957; Petty et al., 1970). The greatest “fatigue” occurs within the first 1-2 minutes. However, SDLB studies (e.g., Petty et al., 1970; Stokinger, Cooper, Meissner, & Jones, 1972) and monaural studies (e.g., Mirabella, Taub, & Teichner, 1967; Wiley, Small, & Lilly, 1973) imply that the “fatigued” ear does not fatigue when the comparison ear is in quiet. What, then, is “fatigue” in SDLB? This paper presents and validates a new model.

The physiology of “fatigue”: the olivocochlear bundle (OCB) in SDLB

An ear’s contribution to loudness is presumed to rise with (1) the number of primary “afferent” neurons (those carrying signals brain-wards) which are firing above their spontaneous rates, and (2) their firing rates (summed in Nizami & Schneider, 1997). An ongoing tone at one ear evokes simultaneous firing (for at least 10 minutes, with a slight firing rate decline) in the OCB of “efferent” neurons (those carrying signals “away from” the brain, periphery-wards) which project to the *opposite* ear, effectively “turning down” that opposite ear’s “volume” as if same-frequency tones there had dropped as much as 24 dB (even more may be possible). Olivocochlear efferents are found at all characteristic (i.e., most sensitive) frequencies of primary afferents, showing a variety of thresholds, allowing smooth and progressive suppression.

Figure 2 illustrates OCB involvement in SDLB, as follows. Contributions to loudness from each ear add with equal weight to create the overall loudness. Subjects equate the contributions by adjusting the “control” (contralateral) ear stimulus intensity during SDLB adjustment sessions. In Fig. 2, at the bottom, is a linear time scale for all of Fig. 2. The figure’s upper and middle frames respectively show the ipsilateral and contralateral ears’ equated contributions to loudness. Stimulus absence is taken as zero intensity. Between contralateral-stimulus presentations, loudness is due only to the ipsilateral stimulus, and does not diminish. The gaps between contralateral stimuli allow the ipsilateral contribution to recover from any contralateral-evoked reduction. The figure’s bottom frame indicates the *average* stimulus intensity at the contralateral ear, “average” because the subject adjusts intensity up and down during the contralateral stimulus’ comparatively brief appearances.

Each ear accesses its separate OCB; stimulus at an ear induces efferent firing which affects the opposite ear. In SDLB, the perstimulatory ipsilateral stimulus progressively “turns down the volume” at the contralateral ear. To compensate, the initial magnitude of the contralateral stimulus intensity must be set increasingly *higher* over successive adjustment sessions (Fig. 2). In response, the *ipsilateral* ear desensitizes, momentarily reducing its contribution to loudness. By the end of each adjustment session, the subject must match that reduced contribution, by reducing the average stimulus intensity from its initial peak to a final steady setting. Typical adjustment sessions of 10 sec (Hood, 1950) to 20 sec (Egan, 1955) are plenty to allow changes in the degree of “volume turn-down” by the OCB, whose initiation has time constants in the hundred-millisecond range.

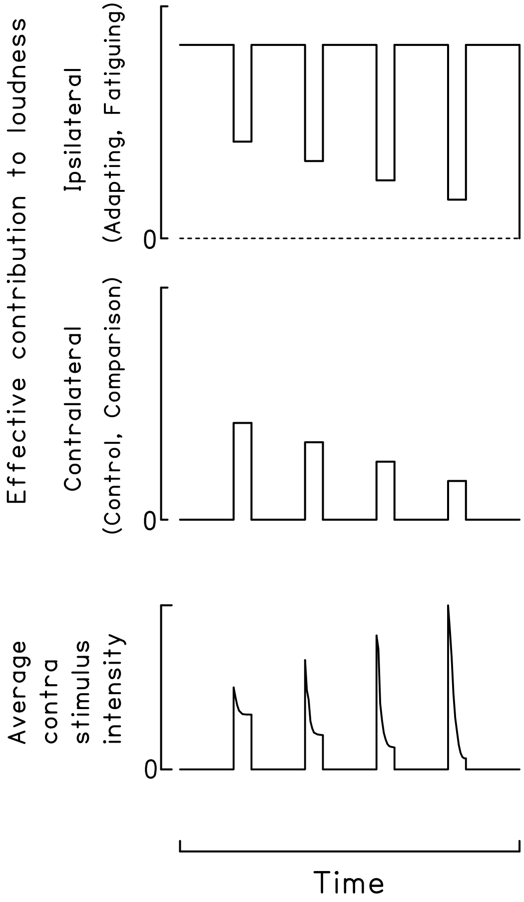


Figure 2. Model of events during the perstimulatory stage of an SDLB run (see text).

Predictions of the model, and evidence consistent with them

Various predictions emanate from the present model, regarding:

(1) *SDLB experiments whose loudness-matching method mimics the changes occurring when subjects make attenuator adjustments*. "Fatigue" behavior should mimic that found in the classic SDLB experiments which involved subjects making attenuator adjustments over 10-20 sec. *Confirmed*: Some experiments from the late 1960s onwards used "the method of constant stimuli", in which the "comparison" tones were made as brief as possible in the belief that "self-fatiguing" by contralateral stimuli could be avoided. See for example Stokinger and Studebaker (1968), Petty et al. (1970, Fig. 4), Stokinger, Cooper, and Meissner (1972), Stokinger, Cooper, Meissner, and Jones (1972), Bray, Dirks, and Morgan (1973), and Dirks, Morgan, and Bray (1974), whose comparison-stimulus durations were respectively 1 sec, 1 sec, 0.2 sec (or 2 sec or 1 sec), 0.2 sec, 3 sec, or 0.3 sec. Of those investigators, only Bray et al. (1973) had subjects perform traditional attenuator adjustments (3 sec), which proved difficult. Such briefness of comparison stimuli does not allow subjects enough time to adjust intensity. Hence, the comparison stimulus was kept at a fixed intensity on any single

presentation, the subject signaling whether it was louder or not than the "fatiguing" stimulus, the experimenter then adjusting its intensity in order to cross back and forth, on a series of successive judgments, the intensity that putatively provided equal loudness. Each such determination would be followed by a rest period for the subject and experimenter while the "fatiguing" stimulus continued to play to the subject.

(2) *Post-stimulatory recovery*. Post-stimulatory, the phenomena of Fig. 2 will reverse, as will "fatigue". *Confirmed*: Carterette (1955, Fig. 3); Egan (1955, Fig. 2); Thwing (1955, Fig. 3).

(3) *The "duty cycles" ([stimulus duration] divided by [stimulus duration plus recovery interval]) of the "fatiguing" and comparison stimuli*. The lower the "duty cycle" of an intermittent squarely-amplitude-modulated stimulus, the more the time for the opposing ear to recover from OCB-mediated "volume turn-down". Thus, an intermittent ipsilateral stimulus should have less accumulated effect on the contralateral ear than a steady one, such that, in return, perstimulatory ipsilateral "fatigue" at any time should be less than for a steady ipsilateral tone, but should increase with duty cycle. *Confirmed*: Carterette (1955). Conversely, a *contralateral* duty-cycle increase "turns down the volume" at the ipsilateral ear, reducing that ear's effect upon the contralateral ear. If duty cycle is *identical* at both ears, the contralateral ear (when its stimulus is absent in-between matches) will be the less influential one. *Confirmed*: Sergeant and Harris (1963); Stokinger, Cooper, and Meissner (1972). (3a) *Contralateral stimulus duration*: (3aa) *Stimuli long enough to be continuously attenuated by the subject*. Shortening the adjustment session duration (e.g., from 20 sec to 10 sec) rushes the subject, who exaggerates the initial, peak contralateral stimulus intensity, which exaggerates "volume turn-down" at the ipsilateral ear, necessitating a lower matching intensity. *Confirmed*: Small and Minifie (1961); compare Hood (1950, Fig. 15) to Thwing (1955, Figs. 3, 4) to Egan (1955, Table IV) for "fatiguing" by 1 kHz at 80 dB SPL. (3aaa) If *perstimulatory ipsilateral stimulus intensity is increased* between runs, the subject follows the same train of adjustments, with the same effects. *Confirmed*: Hood (1950, Fig. 15); Jerger (1957); Stokinger and Studebaker (1968); Petty et al. (1970). (3ab) *Contralateral stimuli too short to allow continuous attenuation by the subject*. The briefer such stimuli, the less time for "volume turn-down" at the ipsilateral ear, hence the higher the matching contralateral intensity. *Confirmed*: Stokinger, Cooper, and Meissner (1972, "Experiment 1", 200 ms tones vs. 2 sec tones); Stokinger, Cooper, Meissner, and Jones (1972). (3ac) *Continuous perstimulatory contralateral stimulus*. Such (with contralateral matching sessions still done intermittently) allows the contralateral ear to continuously "turn down the volume" at the ipsilateral ear. The latter's contribution to loudness hence diminishes, during which the continuous ipsilateral stimulus nonetheless "turns the volume down" at the *contralateral* ear. That effect, too, diminishes over time, thanks to the aforementioned "volume turn-down" at the *ipsilateral* ear. During adjustments, then, the subject's initial resetting of the ongoing contralateral stimulus intensity will not be as high – and the final resetting will not be as low – as for an *intermittent* contralateral stimulus, hence less ipsilateral "fatigue", if any. *Confirmed*: Small and Minifie (1961, Fig. 3). (3aca) *"Fatigue" during the* pre*stimulatory period* must obey the same principles, albeit involving far less accumulated "volume turn-down" at each ear. *Confirmed*: Egan (1955, Fig. 6), Fraser et al. (1970, Table 1), and Petty et al. (1970, Table 1) show contralateral stimuli equated in intensity to same-frequency ipsilateral stimuli.

(4) Momentarily dropping (not to zero) the ipsilateral perstimulatory intensity during the adjustment session. This, combined with the usual manner of setting the initial peak contralateral stimulus intensity (Fig. 2), will produce an even lower matching contralateral loudness contribution (i.e., greater "fatigue"). *Confirmed*: Egan (1955, p. 115 with Fig. 4).

(5) Waveform frequency of ipsilateral and contralateral tones. OCB efferents have V-shaped "tuning curves" of the threshold for stimulated firing versus the stimulus frequency, like those of primary afferents. Hence "fatigue", *as measured momentarily using tones of the same*

frequency to each ear during each adjustment session, should progressively decrease as the contralateral-tone frequency diverges from the (otherwise fixed) frequency of the ipsilateral tone, according to an inverted "tuning curve". *Confirmed*: Thwing (1955); Fraser et al. (1970); Bray et al. (1973). When *ipsilateral and contralateral tones have the same frequency*, "fatigue" should be greatest for frequencies for which the OCB innervation at the organ of Corti is densest, namely, mid-to-high frequencies. *Confirmed*: Jerger (1957).

(6) Presentation of contralateral stimulus after "fatiguing" stimulus. Any contralateral-ear influence on the ipsilateral ear is now irrelevant; subjects hence equate stimulus intensities at the two ears. *Confirmed*: Egan and Thwing (1955); Petty et al. (1970); Fraser et al. (1970).

References

Bray, D.A., Dirks, D.D., Morgan, D.E., 1973. Perstimulatory loudness adaptation. J. Acoust. Soc. Am. 53, 1544-1548.

Carterette, E.C., 1955. Perstimulatory auditory fatigue for continuous and interrupted noise. J. Acoust. Soc. Am. 27, 103-111.

Dirks, D.D., Morgan, D.E., Bray, D.A., 1974. Perstimulatory loudness adaptation in selected cochlear impaired and masked normal listeners. J. Acoust. Soc. Am. 56, 554-561.

Egan, J.P., 1955. Perstimulatory fatigue as measured by heterophonic loudness balances. J. Acoust. Soc. Am. 27, 111-120.

Egan, J.P., Thwing, E.J., 1955. Further studies on perstimulatory fatigue. J. Acoust. Soc. Am. 27, 1225-1226.

Fraser, W.D., Petty, J.W., Elliott, D.N., 1970. Adaptation: central or peripheral? J. Acoust. Soc. Am. 47, 1016-1021.

Hood, J.D., 1950. Per-stimulatory fatigue. Acta Oto-Laryngol. Supp. 92, 26-57.

Jerger, J.F., 1957. Auditory adaptation. J. Acoust. Soc. Am. 29, 357-363.

Mirabella, A., Taub, H., Teichner, W.H., 1967. Adaptation of loudness to monaural stimulation. J. Gen. Psych. 76, 251-273.

Nizami, L., Schneider, B., 1997. Auditory dynamic range derived from the mean rate-intensity function in the cat. Math. Biosci. 141, 1-28.

Petty, J.W., Fraser, W.D., Elliott, D.N., 1970. Adaptation and loudness decrement: a reconsideration. J. Acoust. Soc. Am. 47, 1074-1082.

Sergeant, R.L., Harris, J.D., 1963. The relation of perstimulatory adaptation to other short-term threshold-shifting mechanisms. J. Speech. Hear. Res. 6, 27-39.

Small, A.M. Jr., 1963. Auditory adaptation. In Jerger, J. (Ed.), Modern Developments in Audiology. Academic Press, New York, NY.

Small, A.M. Jr., Minifie, F.D., 1961. Effect of matching time on perstimulatory adaptation. J. Acoust. Soc. Am. 33, 1028-1033.

Stokinger, T.E., Cooper, W.A. Jr., Meissner, W.A., 1972. Influence of binaural interaction on the measurement of perstimulatory loudness adaptation. J. Acoust. Soc. Am. 51, 602-607.

Stokinger, T.E., Cooper, W.A. Jr., Meissner, W.A., Jones, K.O., 1972. Intensity, frequency, and duration effects in the measurement of monaural perstimulatory loudness adaptation. J. Acoust. Soc. Am. 51, 608-616.

Stokinger, T.E., Studebaker, G.A., 1968. Measurement of perstimulatory loudness adaptation. J. Acoust. Soc. Am. 44, 250-256.

Thwing, E.J., 1955. Spread of perstimulatory fatigue of a pure tone to neighboring frequencies. J. Acoust. Soc. Am. 27, 741-748.

Wiley, T.L., Small, A.M. Jr., Lilly, D.J., 1973. Monaural loudness adaptation. J. Acoust. Soc. Am. 53, 1051-1055.

THE HUMAN COCHLEAR MECHANICAL NONLINEARITY INFERRED THROUGH THE SCHAIRER ET AL. (2003) MODEL

Lance Nizami
Independent Research Scholar, Palo Alto, California, USA
nizamii2@att.net

Abstract

Schairer et al. (2003) hypothesized that a multiplicative internal sensory noise, combined with the cochlear mechanical nonlinearity, causes a probe's detection threshold under forward-masking to systematically determine its psychometric function's slope. Here, psychometric functions of unprecedented precision are shown for forward-masked probe-tone detection as a function of intensity of same-frequency forward-masker at fixed masker-probe time gap. Within the Schairer et al. model, these psychometric functions imply that the cochlear nonlinearity's rate-of-change declines as a power function of dB SPL. Rates-of-change, once integrated, give the hypothetical nonlinearity itself, as a function of a single unknown parameter for which suitable values are inferred by comparing hypothetical rates-of-change in man to actual rates-of-change in animals. The model cochlear mechanical nonlinearity in man has similar magnitude and shape to those in animals.

Schairer et al. (2003) introduced a model of the hypothetical influence of the cochlear compressive mechanical nonlinearity upon the slopes of psychometric functions for probe detection under forward-masking. The present paper presents psychometric functions for detection of forward-masked probe tones, and examines whether the slopes of those functions support the Schairer et al. (2003) model. The psychometric functions have a distinct advantage over others, in that they are probably the most precise ever obtained, the inferred probe-detection thresholds having 95% confidence intervals of <2 dB in most cases, and of <1 dB in some cases.

Schairer et al. (2003) imagined that when the cochlear nonlinearity is plotted in scales of dB of output versus dB SPL of input, it forms two line segments, conjoined sharply at a point. Schairer et al. (2003) next noted that the detection of any probe stimulus is empirically characterized by a psychometric function, which, they assumed, exists not because of the nonlinearity per se, but due to the effect of distributions of internal noise upon the output of the nonlinearity in response to input. That is, internal noise was assumed to be multiplicative, thereby having the same distribution at any point along a logarithmic output scale, such as a decibel scale. *For any probe-detection threshold, therefore, which represents some agreed-upon point on a psychometric function, the psychometric function's span would correspond to a constant decibel range of output*. Figure 1 shows the Schairer et al. (2003) model.

The cochlear input-output response is hypothetically linear when the probe's intensity is low, as for weaker forward-maskers (left side of Fig. 1). Then, a given number of decibels of cochlear output will hypothetically correspond to the same given number of decibels SPL, i.e., the psychometric function will have a constant width. The cochlear input-output response is hypothetically compressive (but of constant slope) at moderate threshold probe intensities (right side of Fig. 1). Then, a larger number of decibels SPL will be required in order to span the same given range of decibels of output as before – that is, the psychometric function for forward-masked probe detection will hypothetically have a greater, but constant, width.

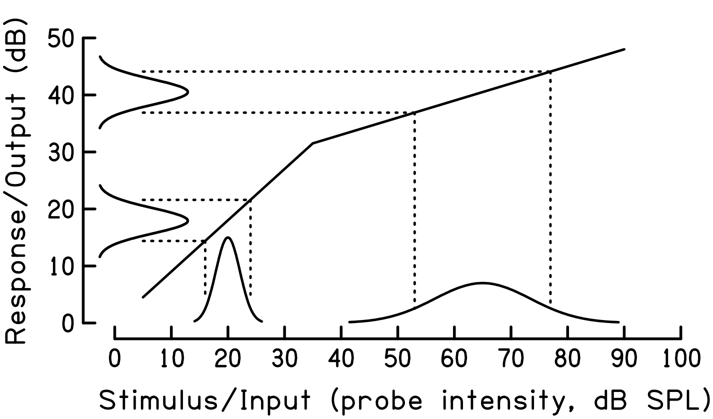


Figure 1. The Schairer et al. (2003) model of the effect of the cochlear mechanical nonlinearity upon the slopes of the psychometric functions for forward-masked probe detection. A pure tone of given intensity evokes a fixed internal response. Internal noise adds to that response, creating a probabilistic distribution of internal responses to a given pure tone (vertical axis). The distribution is assumed to be Gaussian (Green & Swets, 1988). The internal noise is presumably multiplicative, which makes the tone-evoked response distributions on the vertical axis identical, in a decibel scale of internal response, for any tone intensity. Hypothetically, then, the system is tantamount to one being internally noiseless but stimulated by a pure tone whose intensity over repeated presentations follows a Gaussian distribution in a decibel intensity scale (horizontal axis). The mean value of that Gaussian is the detection threshold for the tone (modified from Schairer et al., 2003). Note well that the true amplitudes of the probability density functions are not dB or dB SPL, as might appear from the graph, but rather are probability density, imagined as the label of a z-axis rising perpendicularly out of the page from {0,0}. As such, the shown probability density functions are projections upon the input/output plane of the graph. Also, for illustration's sake, the input distributions shown here are at least twice as wide as will be eventually implied from empirical psychometric functions.

The wider the psychometric function, the shallower it's slope. Within the Schairer et al. (2003) model of the cochlear nonlinearity as two line segments, the psychometric functions for forward-masked probe-detection should therefore have just two possible slopes. But those slopes cannot be known until the detection thresholds are actually established, because two elements of the Schairer et al. (2003) model - the human cochlear input-output response, and the level of the hypothesized internal noise - are unknown. Altogether, then, testing the Schairer et al. (2003) model requires reliably documenting the slopes of the psychometric functions for probe detection, over a broad range of probe-detection thresholds.

Psychometric functions to test the Schairer et al. (2003) model

One way to provide a broad range of probe-detection thresholds is to strongly forward-mask a probe tone, so that its detection threshold will be highly elevated at very short time-gaps between the constant forward-masker and the probe. Here, the probe was a 2 kHz tone having a Gaussian envelope with a standard deviation of 0.5 ms, equal to the tone's period. The forward-masker was a 97 dB SPL 200-ms (not including ramps) 2-kHz tone. Each probe-detection threshold was found using blocks of 100 self-paced two-interval two-alternative forced choices (2I2AFC), during which the forward-masker and probe intensities remained constant (method of constant levels). A double-walled, single-seat soundproof chamber was used, and testing took sufficiently long that only two male adults participated, but with extensive practice. Actual detection thresholds were estimated through Probit Analysis (Finney, 1971), in which the subject's scores (out of 100) are fitted to a cumulative Gaussian, an ogive that is taken to be the psychometric function. It is the integral of an underlying Gaussian probability density function (Gaussian distribution), and is characterized by two numbers inherent to that distribution, namely (1) its mean value in dB SPL, which corresponds to the midpoint of the psychometric function, at which the psychometric function's slope is evaluated, and (2) its standard deviation, which is inversely proportional to the psychometric function's slope.

The Schairer et al. (2003) model in the context of the experiment

According to Schairer et al. (2003), the slope of the psychometric function should take on just two values, one for low probe-detection thresholds, and one for moderate probe-detection thresholds. (High probe-detection thresholds are beyond the scope of most experiments, and were therefore absent from the model.) But the Schairer et al. (2003) nonlinearity is the simplest one imaginable, and a more sophisticated model might posit a nonlinearity that resembles those recorded from animals – an increasingly compressive one, i.e., one whose slope declines monotonically with increasing probe intensity.

In the present experiment, the masker-probe time-gap was fixed at 3 ms, just beyond the range of physical overlap of forward-masker and probe. With increase in forward-masker intensity, the probe-detection threshold rises monotonically, as generally seen in the literature and as found by Schairer et al. (2003, Figs. 2 & 6) and by Schairer et al. (2008, Fig. 2). Also, the psychometric functions generally widen, as found by Schairer et al. (2003, Figs. 3 & 8 [slopes]) and by Schairer et al. (2008, Fig. 3 and Fig. 4 [slopes]). Figure 2 shows the empirical psychometric functions. Subject 1 had more time than Subject 2, who did not experience the 35, 45, 65, 80-, or 90 dB SPL forward-maskers. Generally, the psychometric-function slope decreases with increase in probe-detection threshold, and for either subject, the decrease is adequately fitted by power functions. Figure 3 shows the fits. Hence, if psychometric-function slope is indeed determined by a multiplicative internal noise and the cochlear nonlinearity (Schairer et al., 2003), then the slope of the nonlinearity itself decelerates with increasing sound-pressure-level over roughly 20-80 dB SPL.

The human cochlear nonlinearity by extension of the Schairer et al. (2003) model

The Schairer et al. (2003) model was extended here in order to reveal the human cochlear nonlinearity. The extension starts with a key assumption: that the Gaussian-shaped "input distribution" of the model is, in fact, the same Gaussian probability density function which can be integrated to make the psychometric function for the detection of the forward-masked probe. Two realizations were also required, viz., that (1) the average slope of the cochlear nonlinearity itself can be measured between any two points on said nonlinearity, and (2) that those two points can correspond to the "edges" of a (symmetric) psychometric function which is centered midway between the two points on the nonlinearity. The psychometric-function width therefore forms the denominator of the average slope of the nonlinearity; the numerator, according to the Schairer et al. (2003) model, is some unknown, but fixed, number of decibels of output. Figure 4 shows the extension of the Schairer et al. (2003) model.

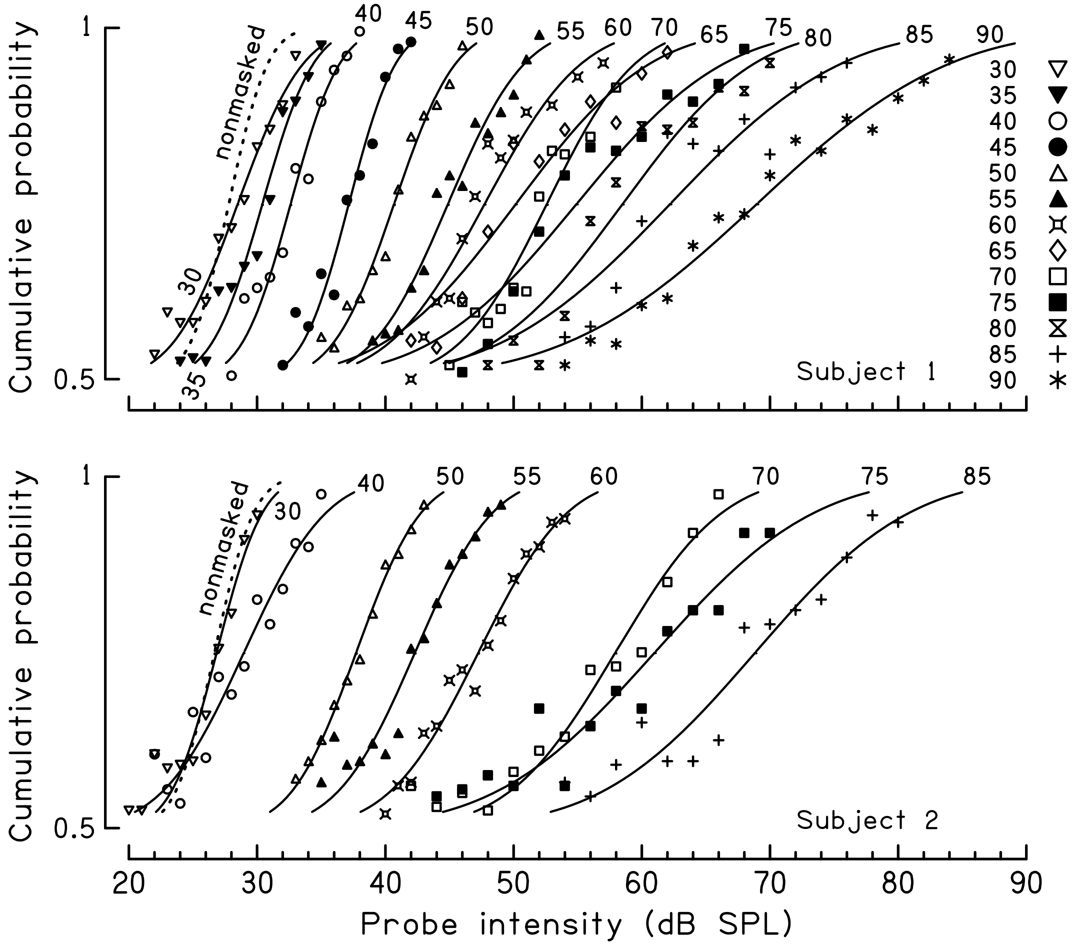


Figure 2. Psychometric functions for forward-masked probe detection and the percentages-correct on which they were based. The two columns on the right-hand-edge of the upper frame match forward-masker intensity to data-plotting symbol.

The widths of the obtained psychometric functions are inversely proportional to their slope. Altogether, then, the average slope of the cochlear nonlinearity over some interval centered on a particular intensity is directly proportional to the slope of the psychometric function for forward-masked probe detection whose centroid corresponds to that intensity. Therefore, quantifying psychometric-function slope as a function of intensity leads to a further equation, in one unknown multiplicative parameter, for the average slope of the cochlear nonlinearity with intensity. Plotting the latter on the same graph as empirical animal-derived curves of the nonlinearity slope allows comparisons which suggest appropriate values of the unknown parameter. To do so, however, the animal-derived curves must be shifted to higher SPLs, because the present probe is much shorter in duration and lower in frequency than the probe tones used in animals, and hence, overall, has much less energy.

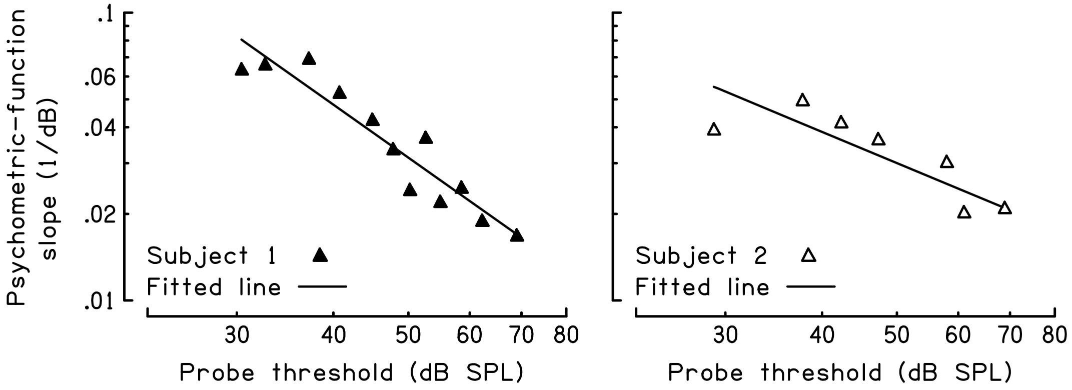


Figure 3. The slopes of the psychometric functions for probe detection (the functions of Fig. 2), versus the actual probe-detection thresholds. The straight lines are fitted power functions.

Equations for the inferred average slope of the cochlear nonlinearity can be integrated to give the nonlinearity itself. The integrals have a lower limit, which is the starting point of the nonlinearity, here assumed to be the probe's detection threshold in the absence of the forward-masker. The cochlear nonlinearities predicted from the experimental results resemble animal recordings, in that they show no distinct point of bending. The upper slopes of the inferred nonlinearities are similar to that of the Schairer et al. (2003) model. The range (in decibels) from maximum to minimum output of the inferred nonlinearity is of the same order of magnitude as those seen in animals. Figure 5 shows the inferred cochlear mechanical nonlinearities.

Finally, it is conceivable that the value of the unknown parameter in the present model could change from tone frequency to tone frequency within a single subject, and could change from subject to subject for a given tone frequency.

Acknowledgments. Data presented here was previously unpublished from Nizami (1999). This paper funded by the author. My thanks to Claire S. Barnes (VAPAMC) for proofreading.

References

Finney, D.J. (1971). Probit Analysis. London: Cambridge University Press.

Green, D.M. & Swets, J.A. (1988). Signal Detection Theory and Psychophysics. Los Altos, California, USA: Peninsula Publishing.

Nizami, L. (1999). On Auditory Dynamic Range. Doctoral dissertation, School of Graduate Studies, University of Toronto, Toronto, Ontario, Canada.

Rhode, W.S. & Recio, A. (2000). Study of mechanical motions in the basal region of the chinchilla cochlea. Journal of the Acoustical Society of America, 107, 3317-3332.

Schairer, K.S., Messersmith, J., & Jesteadt, W. (2008). Use of psychometric-function slopes for forward-masked tones to investigate cochlear nonlinearity. Journal of the Acoustical Society of America, 124, 2196-2215.

Schairer, K.S., Nizami, L., Reimer, J.F. & Jesteadt, W. (2003). Effects of peripheral nonlinearity on psychometric functions for forward-masked tones. Journal of the Acoustical Society of America, 113, 1560-1573.

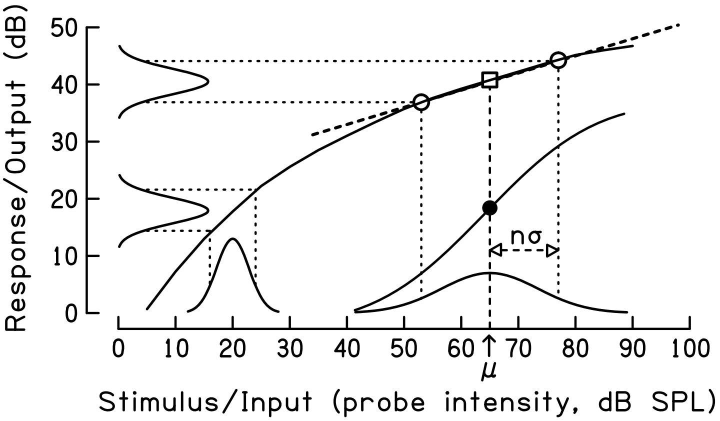


Figure 4. The average slope of the cochlear nonlinearity vs. the width of the psychometric function. Fig. 1 is modified such that the hypothetical cochlear nonlinearity is smoothly-changing (after chinchilla cb24 of Rhode and Recio, 2000). Each of the input distributions is now presumed to be integrated to yield the probe-detection psychometric function, which runs from 0.5 to 1 in the (2I2AFC) experiment. For the right-hand input distribution here, the mean value (and probe-detection threshold) is μ, coinciding with the centroid of the psychometric function (solid dot). The open square is the corresponding locus on the cochlear nonlinearity. The width of the psychometric function is defined as $2n\sigma$, where $n \in {}^{+}\mathrm{I}$; its corresponding points on the nonlinearity are marked by the open circles. Through those circles passes the dashed slanted line, whose slope is the average of the slopes between the two open circles, approximating the nonlinearity's slope at the open square. The approximate slopes are better than apparent, as the input distributions (and corresponding psychometric functions) shown here are (as in Fig. 1) at least twice as wide as implied from the empirical psychometric functions. The psychometric function does not have units of dB, but rather percentage correct, and as such is a projection upon the plane of the graph.

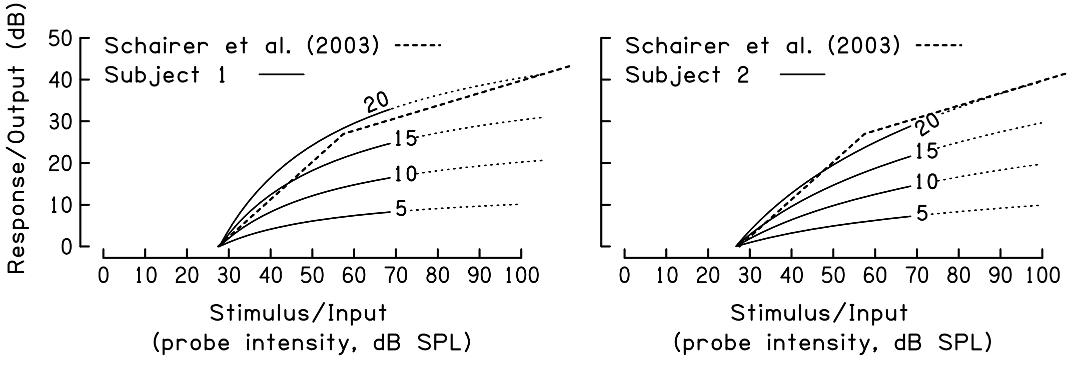


Figure 5. The inferred cochlear nonlinearity (solid lines). The plot labels are suitable values of the one unknown parameter, inferred from cochlear nonlinearities in animals. The dashed lines show the hypothetical nonlinearity of Schairer et al. (2003), adjusted to start at the same point as the solid lines, which are made to originate at an "output" of 0 dB and at the subject's absolute probe-detection threshold. The dotted lines extrapolate.

JAMES McKEEN CATTELL AND THE METHOD OF CONSTANT STIMULI IN THE PSYCHOPHYSICS OF MOVEMENT

Roger Adams[1], Kwee-Yum Lee[2], Gordon Waddington[3] and Hae-Jung Lee[4]
[1]University of Sydney, [2]Australian Catholic University, [3]University of Canberra, [4]Silla University
Faculty of Health Sciences, University of Sydney, Australia
<roger.adams@sydney.edu.au>

Abstract

In 1892, Cattell & Fullerton published a paper on "The Psychophysics of Movement" in which they summarised their findings from the application of Fechner's methods to the study of the extent, force, and time of arm movements made without the aid of vision. Using a form of Munsterberg's apparatus (Titchener, 1905) they found the method of right and wrong cases to be 'the most accurate of the methods' and the difference associated with 75% correct judgement to be the 'most convenient measure of discrimination'. By 1990s, however a major work on examining the role of proprioception in joint stability listed only the methods of adjustment (Joint Position Sense) and limits (Kinesthesia) as being ways of measuring movement sensitivity. Recent technical developments have enabled researchers to employ the Cattell & Fullerton method to obtain psychophysical measures of proprioception derived from the comparison of active movements made to physical stops.

The first experiments on the psychophysics of active movement were conducted by Fechner (1860) with judgments of the amount of force required from the upper limb for it to overcome the resistance due to gravity for a specific weight. The judgments were then used to derive a measure of sensitivity. The experimental method employed by Peirce and Jastrow in 1884 involved judgment of the finger force needed to counter an upwards pressure on beam of a post-office scale from a weight placed in the pan, but it was not until the work of Fullerton & Cattell (1892) that the first judgments were made of differences in the extents of active movements made to physical stops.

Methods of Constant Stimuli, Adjustment, and Limits in the Psychophysics of Movement

The use of comparison of the extent of movements made to physical stops, without the aid of vision, as an experimental method for research into the psychophysics of movement seems to have been jointly developed by James McKeen Cattell and Hugo Munsterberg, who had been classmates in Wundt's Psychophysical Seminar series in Leipzig in the summer of 1885 (Blumenthal, 1997). Munsterberg later supervised the dissertation of Edmund Burke Delabarre, presented in Freiburg in 1891, in which an apparatus he devised (Figure.1) for assessing both horizontal and vertical arm movements was used (Worringham, 1992). Edward Titchener (1905) lists his Fig. 30 as 'Munsterberg's apparatus' and notes that the car, which travels on three horizontal tracks, has a vertically-set brass cylinder for the subject's forefinger, and 'two sliding blocks, which can be set at any point along the middle track, to mark the beginning and end of the two movements' (p.104).

In contrast, the apparatus used by Fullerton and Cattell (Figure 2) had a horizontal brass ring between the front and back wheels of the carriage, into which the seated subject inserted their finger to make a movement. The carriage ran in grooves cut into a brass plate fixed to a tabletop, and metal pins could be inserted to define the extents of movements. A screen obscured vision.

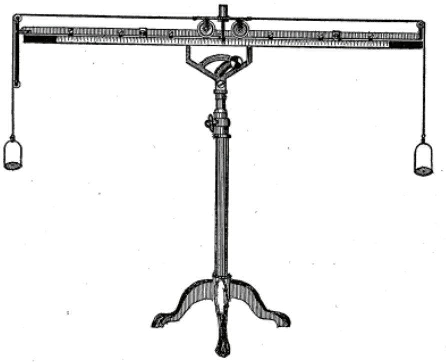


FIG. 30.

Figure 1. Munsterberg's apparatus for the comparison of movements of the arm. From Titchener (1905).

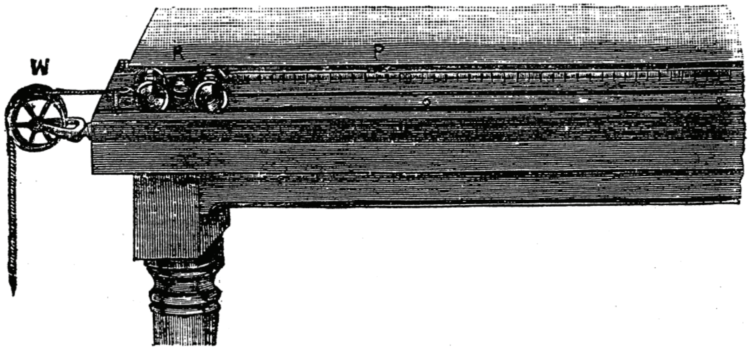


FIG. 2.—APPARATUS FOR MEASURING THE EXTENT OF MOVEMENT.

Figure 2. Apparatus for measuring discrimination of the extent of arm movements. From Fullerton & Cattell (1892).

The adjustable stops for limiting movement of the trolley and defining the beginning and end of arm movements, S' and S'' can be seen more clearly in the apparatus used by Charles Myers (1911) at Cambridge University (Figure 3).

Cattell's first doctoral student at Columbia University, Robert Sessions Woodworth, later supervised the dissertation of Harry Hollingworth, who substituted wood-fiber wheels in place of the original metal ones, to 'more completely eliminate the noise made by the moving carriage' (Hollingworth, 1909, p.7). However, Hollingworth only used the Adjustment Method (Average Error) in his studies, and titled his dissertation 'The Inaccuracy of Movement'. His somewhat pessimistic conclusion to his work was that 'great uncertainty arises in the application of the psychophysical methods to the study of movements' (p.82). Hollingworth's research marked the end of the early study of movement psychophysics.

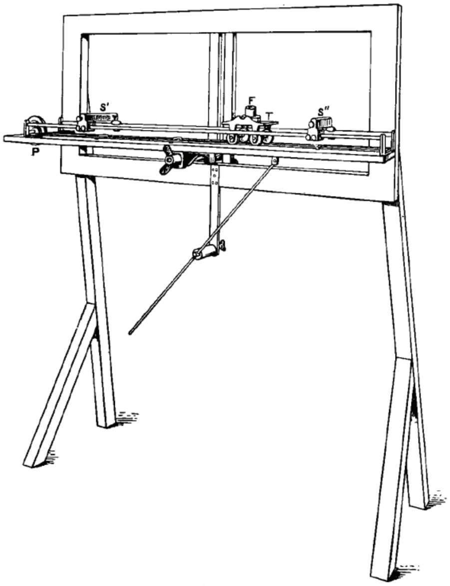


FIG. 5.

Figure 3. Apparatus for studying the limen of just-perceptible active movement. From Myers (1911).

PLATE II

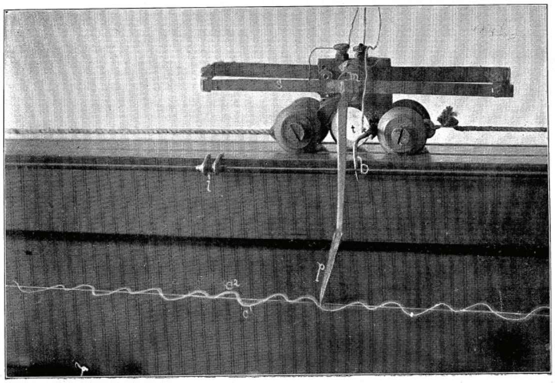


Figure 4. Movable carriage for studying arm movement. From Hollingworth (1909).

The Use of Stopped Movements in Movement Psychophysics

Fullerton & Cattell (1892) noted that it is only in the method of right and wrong cases (later constant stimuli) that both movements ‘are determined by two uprights’, whereas in the adjustment methods the movable upright at the end of the movement is sometimes removed. The inference here is that tasks asking subjects to compare different kinds of movements will involve cognitive processes not usually associated with everyday movement control. Indeed, Laszlo (1992) has argued that when the first (or criterion) movement is made to a stop, and the subsequent reproduction movement is ended in space by the subject, without a stop, then the assumption that the two movements rely on the same sensory-motor processes is not a valid one.

Early in the twentieth century, research into the psychophysics of movement had become directed towards questions associated with simultaneously controlling or manipulating the duration, force and extent of movements, the psychophysical method that was to be used, the time order error, and whether subjects should be permitted to use ‘equal’ or ‘doubtful’ judgments. Others, however, saw the discrimination sensitivity index as a useful measure, separate from the question of psychophysical scaling.

The Einstellung Psychophysical Judgment Process for Movement

Georg Elias Muller, the German researcher to whom Fechner gave his set of weights, proposed a different basis for comparison underlying psychophysical judgment than Fechner. Whereas Fechner argued that in the method of right and wrong cases each lifted weight was compared with a memory reference value that was the average of the two weights (Link, 1992), Muller in 1889 proposed the ‘Einstellung’ hypothesis. By this account, on the second lift, the subject unconsciously sets the same motor command parameters that had previously been successful in overcoming the downwards force on the first lift (Boring, 1929). If the result was that the arm moved up more quickly, the second weight was judged lighter, and if more slowly, judged heavier. Applied to stopped movements, this account would suggest that subjects, while attempting to move at a constant pace, generate an expected stop position, and if the movement continues past this, it is judged longer than on the previous trial, and if stopped before it, judged shorter.

Muller was visited in Gottingen by Charles Spearman, who subsequently argued that the most promising application of the psychophysical measure of sensitivity was to be found in describing individual differences in ability and distinguishing between conditions which either enhanced or worsened the ability to discriminate (Spearman, 1908).

Obtaining an index of discrimination sensitivity to compare conditions of testing under experimental manipulation, in the manner suggested by Spearman, however, must address the difficulties clearly outlined by Cattell and Fullerton (1892) when they summarized the work of Fullerton and Cattell (1892). Here they noted that the method of right and wrong cases was the most accurate of Fechner’s methods for studying the psychophysics of movement, but that it required a considerable number of trials and “is consequently not well suited for provisional, anthropometric, or clinical purposes” (p.447). As a measure, they suggested that “the probable error, that is the difference with which an observer is right 75% of the time, is the most convenient measure of discrimination”, calculated using the probability integral.

Psychophysical Measures in Proprioception Testing and Joint Injury Assessment

A domain with an obvious need for accurate and valid assessment of sensitivity is the clinical musculoskeletal domain, especially with respect to assessment of athletes following joint injury. It was therefore possibly the time cost of obtaining probable error estimates that meant that 100 years later, only two methods of obtaining a measure of sensitivity were listed for testing joint proprioception (Lephart, Riemann, & Fu, 2000), these being versions of the Method of Average Error (Adjustment) and the Method of Limits, and listed as Joint Position Sense testing, and Kinesthesia testing, respectively.

Because the ankle is the most frequently injured joint in the body (Witchalls et al., 2012), an accurate measure of movement discrimination at the joint is needed for assessing the extent of injury to an athlete and for success of rehabilitation (Figure 5). Cattell’s method of using comparison of active movements made to physical stops without the aid of vision therefore needs to be adapted to testing movements in lower limb rather than upper limb joints, and needs to be made more efficient to reduce the time demand of testing.

Three developments have enabled this. The first was the development of the Method of Single Stimuli (Woodworth & Schlosberg, 1966) which halves the number of trials by removing the standard stimulus. The subject must use a memory representation of the stimulus values to make judgments. The second development was non-parametric signal detection analysis via ROC curves (Swets, 1988) which permits the calculation of a bounded discrimination index, the AUC (Area Under the Curve). Finally, the availability of computer-controlled stepper motors, used to accurately position read heads on hard drives and control spot welding in auto manufacturing, has provided a laboratory system for resetting physical stops that is fast, error-free, and able to replicate stop settings to tolerances of 1/100 of a millimeter.

Figure 5. Active Movement Extent Discrimination Apparatus (AMEDA) for testing sensitivity to differences in extent for ankle plantarflexion and inversion movements. The physical stop blocks labeled ‘D’, ‘M’ and ‘S’ set the Deep, Mid, and Shallow distances down from horizontal, at 14, 11 and 8 degrees respectively. The aluminum plates, in a row at the front of the apparatus, provide a set of depths to discriminate for each block.

With these developments, it has become possible to continue the psychophysical research on stopped movements that Cattell started 120 years ago. Discrimination measures obtained using various versions of the AMEDA (Active Movement Extent Discrimination Apparatus) have been used to quantify the followings: the value of balance board rehabilitation (Waddington et al., 1999); the effects of textured athletic shoe insoles (Waddington & Adams, 2003); the risk of poor hip movement discrimination score for subsequent hamstring injury (Cameron et al., 2003); the role of vision in judging neck movements (Lee et al., 2004); and the effects of retraining on movement discrimination after anterior shoulder dislocation (Naughton et al., 2005). The research sequence involves identifying the action of interest, then designing a system for stopping a set of different movement extents that can then be used to generate a score representing a person’s ability to differentiate extents on the dimension of interest. As there are almost an infinite number of combinations of actions of interest in different individual/contexts, this method seems to have a potential to become one of most promising applications in psychophysics as Spearman foresaw.

References

Boring, E.G. (1929). *A history of experimental psychology*. New York: Appleton Century Crofts.

Blumenthal, A.L. (1997). Wilhelm Wundt. In W.G. Bringmann, H.E. Luck, R. Miller & C.E. Early (Eds.), *A pictorial history of psychology*. Chicago: Quintessence Publishing Co.; pp. 117-125.

Cameron M., Adams, R., & Maher, C. (2003). Motor control and strength as predictors of hamstring injury in elite players of Australian football. *Physical Therapy in Sport*, *4*, 159-166.

Cattell, J.M., & Fullerton, G.S. (1992). The psychophysics of movement. *Mind*, New Series No.3, 447-452.

Fechner, G.T. (1966). *Elements of psychophysics* (Vol. 1) (E.G. Boring & D.H. Howes. Eds.; H.E. Adler, Trans.). New York: Holt, Rinehart & Winston. (Originally published 1860).

Fullerton, G.S., & Cattell, J.M. (1892). *On the perception of small differences: with special reference to the extent, force and time of movement.* Philadelphia: Uni. of Pennsylvania Press.

Hollingworth, H.L. (1909). The inaccuracy of movement. *Archives of Psychology*, *13*, 1-87.

Laszlo, J.I. (1992). Motor control and learning: How far do the experimental tasks restrict our theoretical insight? In J.J. Summers (ed.) *Approaches to the study of motor control and learning*. North-Holland: Elsevier Science Publishers, pp. 47-79.

Lee, H., Nicholson, L.L., & Adams, R.D. (2004). Sensitivity to differences in the extent of neck-retraction and -rotation movements made with and without vision. *Perceptual and Motor Skills*, *98*, 1081-1089.

Lephart, S.M., Riemann, B.L., & Fu, F.H. (2000). Introduction to the sensorimotor system. In S.M. Lephart & F.H. Fu (Eds.), *Proprioception and neuromuscular control in joint stability*. Champaign, IL: Human Kinetics.

Link, S.W. (1992). *The wave theory of difference and similarity*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Myers, C.L. (1911). *A text-book of experimental psychology with laboratory exercises.* Cambridge: Cambridge University Press.

Naughton, J., Adams, R., & Maher, C. (2005). Upper-body wobbleboard training effects on the post-dislocation shoulder. *Physical Therapy in Sport*, *6*, 1-7.

Peirce, C., & Jastrow, J. (1885). On small differences of sensation. *Proceedings of National Academy of Science, 3*, 75-83.

Spearman, C. (1908). The method of 'right and wrong cases' (constant stimuli) without Gauss's formulae. *British Journal of Psychology, 2*, 227-242.

Swets, J.A. (1988). Measuring the accuracy of diagnostic systems. *Science, 240*, 1285-1293.

Titchener, E.B. (1905). *Experimental psychology*. (Vol. II, Part I). New York: Macmillan.

Waddington, G., Adams, R., & Jones, A. (1999). Wobble board (ankle disc) training effects on the discrimination of inversion movements. *Australian Journal of Physiotherapy*, *45*, 95-101.

Waddington, G., & Adams, R. (2003). Football boot insoles and sensitivity to ankle inversion movement extent. *British Journal of Sports Medicine*, *37*, 170-175.

Witchalls, J., Waddington G., Blanch P., & Adams, R. (2012). Ankle instability effects on joint position sense when stepping across the Active Movement Extent Discrimination Apparatus. *Journal of Athletic Training*, in press.

Worringham, C.J. (1992). Some historical roots of phenomena and methods in motor behavior research. In G.E. Stelmach & J.Requin (Eds.), *Tutorials in motor behavior II*. North-Holland Amsterdam: Elsevier Science Publishers.

COMPARISONS OF INDICES OF MOVEMENT DISCRIMINATION: PSYCHOMETRIC FUNCTION, INFORMATION THEORY, AND SIGNAL DETECTION ANALYSIS

Kwee-Yum Lee[1], Roger Adams[2], Hae-Jung Lee[3], and Gordon Waddington[4]
[1]Australian Catholic University, [2]The University of Sydney, [3]Silla University,
[4]University of Canberra
School of Exercise Science, Faculty of Health Sciences, Australian Catholic University
kwee.lee@acu.edu.au

Abstract

Quantitative assessments of movement discrimination ability are used in sport and rehabilitation to gain insights into the mechanisms of neural function, skill performance under different conditions and recovery after injury. Here, we have for the first time compared three psychophysical measures of discrimination of extent of ankle movements. The movement discrimination sensitivity indices were; the just noticeable difference (JND) derived from the psychometric function, the amount of transmitted information (TI) from information theory, and the area under the ROC curve (AUC) from non-parametric signal detection analysis. Absolute judgements were obtained from 25 participants for the extents of active ankle inversion movements made to physical stops at different depths. Only the AUC showed a significant difference in discrimination scores with depth and a downwards linear trend of lesser scores with greater movement extents. These results are consistent with the previous findings of Cattell and Fullerton (1892) and Magill and Parks (1983).

Movement sensitivity, or proprioception, measures are among the first methods used in the psychophysics domain in the late 19th century. Fechner used lifting different weights to gain insight into the relationship between measurable difference in weights and related perception (Fechner, 1966). Continuing in his footsteps, most of the psychophysicists during this early period examined the movement sensitivity question with proprioception tests (Cattell & Fullerton, 1892; Fernberger, 1913; Peirce & Jastrow, 1885; Urban, 1910). However, unanswered questions related to movement sensitivity did not stimulate research for the next 100 years; the topic is mostly only introduced in relation to the early history of psychophysics.

During this period, however, other areas such as neurophysiology, rehabilitation medicine and physiotherapy have utilised proprioceptive sensitivity measures to assess interventions intended to improve motor function, or to describe the mechanisms underlying proprioception and motor control. Most of the methods employed in these fields have used passively applied movements to find a threshold for detection (Refushauge, Kilbreath, & Raymond, 2000) or have used open-ended movements such as reproducing a criterion movement on one limb with a contralateral limb (Willems et al., 2005). Accuracy in matching end position, however, may not reflect the precision of discriminating the differences in position (John, Goodwin, & Darian-Smith, 1989), and there are only a handful of studies in this area utilising the methods of constant stimuli and psychophysical measures. In fact Cattell's statement (1892) that the method of constant stimuli was the most reliable psychophysical measure for assessing movement discrimination was unknown in the domains of allied health science until 1990s.

Recently, some researchers have used the method of constant stimuli to obtain the just noticeable difference (JND) measure to quantify the sensitivity of the arm movement extent, beginning with Carlton & Newell (1985), and Magill & Parks (1983). Meanwhile with the advent of information theory by Shannon in the 1950s (Shannon & Weaver, 1963), some neurophysiologists have employed an information index to quantify proprioception sensitivity, namely transmitted information (TI). For example, Clark and colleagues (1995) used the amount of TI to resolve the maximal range of motion that can be discriminated at the finger. As this measurement method is based on the confusion matrix, the use of constant stimuli can also be applied to obtain the TI value. Waddington, Adams and colleagues have used the method of single stimuli with absolute judgement to create a receiver operating characteristic (ROC) curve and the area under the curve (AUC) for a kinaesthetic sensitivity measure. In particular, this group of researchers have examined sensitivity at a variety of joints (e.g., neck, ankle, leg, and shoulder) with different test conditions to examine effects of joint pain and joint injury (Lee, Nicholson, Adams, & Bae, 2005; Waddington & Adams, 1999).

To date, no studies have compared different sensitivity measures in order to examine the robustness of each method in comparison with the method of constant stimuli. We believe this to be a necessary methodological comparison that is relevant to future research requiring accurate and practical evaluation of proprioception. To answer the question, the task we chose was judgment of extent of ankle inversion movement at three different base depths, performed in an upright posture. Ankle inversion movement was actively produced by the subject to a defined angle. With a variation in the angle of the stimulus presented, and absolute judgement recorded for each trial, it was possible to compare different sensitivity measures – the just noticeable difference JND, transmitted information TI, and the area under the ROC curve, the AUC.

Method

Participants. Twenty five university students volunteered to take part in this study. All of them were healthy with no movement disorders or ankle injury in the 6 weeks preceding the study and were reported to be right foot dominant.

Apparatus. The Active Movement Extent Discrimination Apparatus (AMEDA) was employed for data collection. The AMEDA consisted of two wooden blocks (410 mm × 410 mm × 410 mm from the floor) which provided a platform for a participant to stand on. One wooden board was fixed while the other rotated along the long axis of the foot which could be moved with the inversion movement of the foot until it was stopped by a wooden block placed below on the outer edge of the board. The wooden blocks of three different heights were designated as Shallow (8°), Mid (11°), and Deep (14°). A single block was used for a test series, because with these blocks, different stop-heights could be tested with active inversion of the foot by use of five aluminium spacers with different thickness (1 mm–5 mm) that could be slipped in underneath the wooden block, and by including the 'no spacer' possibility, six different inversion depths were able to be generated for each of the three blocks. The angular depth (S: stimulus) for a corresponding metal spacer and the range for each depth is show in Table 1.

Table 1. Angular depth of each stimulus (S) and the range of each condition.

	S1	S2	S3	S4	S5	S6	Range
Shallow (8°)	8.21	8.55	8.90	9.20	9.54	9.90	1.69
Middle (11°)	10.82	11.16	11.52	11.82	12.16	12.52	1.70
Deep (14°)	13.87	14.21	14.58	14.88	15.23	15.59	1.72

Procedure. The participant stood on the AMEDA with the tested foot on the movable board. Upon a 'go' signal, a steady ankle movement was produced until the edge of the board was stopped by the wooden block. The participant returned to the initial horizontal position and gave a number judgment. Familiarisation trials with 6 different depths were conducted, informing the corresponding number of to each depth.

During trials, participants made foot movements, followed by an immediate reporting of a number between 1 and 6 to indicate the depth felt by the extent of movement. The 48 trials for each block were conducted on the right foot in random order, with a 2-min rest between blocks of trials. Maintenance of an even distribution of weight on both legs was visually monitored by the experimenter throughout the trials.

Analysis. The stimulus value on each trial and the corresponding absolute judgement were collected and used for three different analyses to obtain JND, TI, and AUC.

JND: Percentage of greater than standard stimulus (S3) is plotted against variable stimulus values, and probit analysis (Matlab, Mathwork) was used to find the best fit for the cumulative normal curve. With this curve, the just noticeably less (JNL), the stimulus which is judged to be less than standard on 75 per cent of trials, and the just noticeably greater (JNG), the stimulus which is judged to be greater than standard on 75 per cent of trials were calculated. The average just noticeable difference (JND) is calculated by halving the intervals between JNL and JNG:

$$\mathrm{JND}=\frac{|\mathrm{JNG}-\mathrm{JNL}|}{2}. \tag{1}$$

TI: The amount of presented stimuli, response frequency, and joint occurrence of the stimulus-response match can be quantified in bits of information using a confusion matrix. From this matrix, average amount of transmitted information can be calculated as below:

$$T(x;\, y)=H(x)+H(y)-(x,\, y), \tag{2}$$

where $T(x;y)$ is the amount of information transmitted from stimulus to response. $H(x)$ is the estimated information-per-stimulus, $H(y)$ is the estimated information per response, and $H(x,y)$ is the estimated information in the joint occurrence of a stimulus and response (Attneave, 1959).

AUC: angle stimuli were considered as noise and signal in a pair-wise manner (1-2, 2-3, 3-4, 4-5, 5-6), and response values to each stimulus pair were treated as a rating scale for the signal and noise presentation (McNicol, 2004). For each consecutive stimulus pair, 1-2, 2-3, 3-4, 4-5, and 5-6, signal and noise correspond to greater and lesser values (Welford, 1976). The Matlab (Mathworks) program was used to calculate the ROC curve.

A repeated-measures ANOVA was conducted on each discrimination measure using SPSS for Windows. Polynomial trend contrasts was used to test for the linear trend in scores across three depths (Shallow, Mid, Deep) for each measure.

Results

Mean JND, TI and AUC values for three different depths are shown in Table 2 and Figure 1. JND values ranged from 1.12° to 1.26°, and average information transmitted ranged from 0.48 to 0.56 bits. Both measures did not show any significant linear trend with depth, $F(1, 24) \leq 2.74$, $p \geq .11$. AUC values ranged from 0.56 to 0.59, and the ANOVA showed a significant linear decrement in discrimination ability as the block depth decreased from Shallow to Deep, $F(1, 24) = 5.17$, $p < .05$. Across the stimulus pairs, the AUC also showed a significant downward linear trend, $F(1, 24) = 4.61$, $p < .05$. Hence, according to the AUC measures, discrimination of ankle inversion angle was significantly more difficult at the deeper positions.

Table 2. Mean (± standard deviation) values of discrimination measures of inversion movement in three depths.

	Shallow	Mid	Deep
JND (°)	1.12 ± 0.89	1.20 ± 0.78	1.26 ± 0.80
TI (bits)	0.56 ± 0.25	0.54 ± 0.13	0.48 ± 0.09
AUC	0.59 ± 0.06	0.56 ± 0.04	0.56 ± 0.04

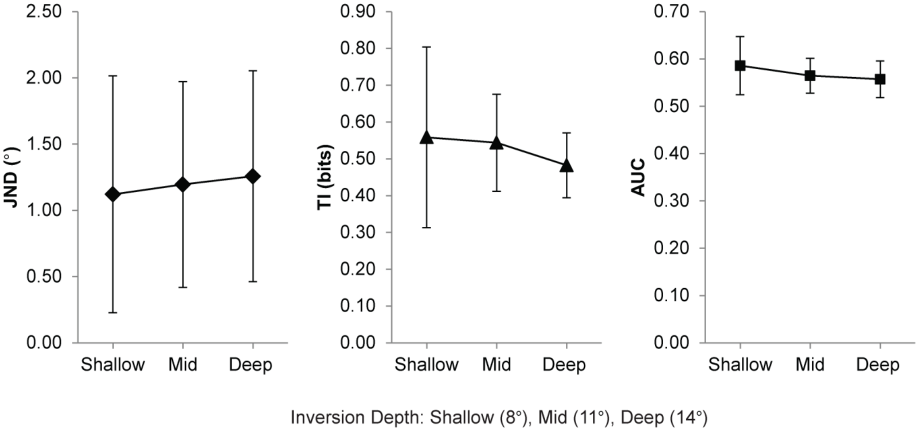


Figure 1. Discrimination measures - JND, TI, and AUC - in Deep Mid and Shallow depth condition. JND: just noticeable difference, TI: transmitted information, and AUC: area under the ROC curve.

Discussion

As the depth increased, mean sensitivity values of all measures tend to decrease, but only AUC measures were able to detect a significant trend of the discrimination ability becoming worse. The significant downward linear trend of AUC measures with increasing movement extent is in line with other previous findings examining the discrimination measures under active (Magill & Parks, 1983) or passive arm movement conditions (Carlton & Newell, 1985), where increase in JND values with movement amplitude were reported. Similarly, in the movement matching task, mid-range of joint angles showed the highest accuracy (Clark et al., 1979). Cattell and Fullerton (1892) also mentioned that the Weber's law does not seem to hold for the perception of movement, but they found rather JND increased proportionally with the square root of the movement extent. All these findings and observations are congruent with the results obtained by AUC measure. Among JND, TI and AUC measures, we might safely assume that AUC method to have the highest sensitivity for measuring movement discrimination ability when using a total of 144 trials of stimulus presentation and the method of absolute judgement.

The number of trials might be one of the reasons why JND did not show a significant difference with depth. In calculating JND, the sensory error of each stimulus is assumed to follow the Gaussian Law of error distribution. To obtain this probability curve, 48 trials for each depth might not have been enough for obtaining the best estimates from the probit analysis. Early studies of JND on movements usually employed around 1,000 trials to obtain reliable estimation of the probability curve (Cattell & Fullerton, 1892; Peirce & Jastrow, 1885). With increased number of trials, we might see similar results between JND and AUC, but the sheer number of trials lacks practicality.

The channel capacity model using the Shannon index is used for the TI measure. This measure represents how much information has been transmitted through the human central nervous system. Our results showed that the discrimination task using method of single stimuli involved around 0.5 bits of transmitted information. In their matching task using proximal interphalangeal joints, Clark and colleagues (1995) found the TI values to be 1.4–1.9 bits. As these studies utilised a total of around 1,000 trials to estimate the probabilities using a confusion matrix, again, the number of trials might have been the limiting factor and the reason why the measure failed to produce significant results.

The AUC measure is calculated using the underlying model of signal plus noise versus noise distribution curves on the sensation continuum. By using presentations of trials and the associated responses, the probability curve for correct detection of the signal, and false alarm rate, are obtained for plotting a ROC curve (Snodgrass, Berger, & Haydon, 1985). The sensitivity of detection represented by this method is the same for one subject regardless of the different criterion used (McNicol, 2004). As the measure can report the expected difference in movement sensitivity with increasing depth, this method is not only sensitive but also practical in its application due to relatively fewer trials involved. Several studies using this method have addressed a wide range of applied questions associated with proprioception sensitivity, such as the issue of limb dominance and proprioception (Cameron & Adams, 2003) or neck proprioception with frequent subclinical neck pain (Lee, Nicholson, Adams, & Bae, 2005). With its practicality and sensitivity, AUC measure will be useful for future studies examining proprioception and motor function by providing a viable tool to quantify movement discrimination ability.

References

Attneave, F. (1959). *Applications of information theory to psychology: A summary of basic concepts, methods, and results.* New York: Holt, Rinehart, and Winston.

Cameron, M., & Adams, R. (2003). Kicking footedness and movement discrimination by elite Australian Rules footballers. *Journal of Science and Medicine in Sport*, 6, 250-258.

Carlton, L. G., & Newell, K. M. (1985). A psychophysical examination of the perception of movement extent under passive movement conditions. *Journal of Human Movement Studies, 11*, 35-47.

Cattell, J.M, & Fullerton, G.S. (1892). On the perception of small differences: With special reference to the extent, force and time of movement. Philadelphia: University of Pennsylvania Press.

Clark, F.J., Horch, K.W., Bach, S.M., & Larson, G.F. (1979). Contributions of cutaneous and joint receptors to static knee-position sense in man. *Journal of Neurophysiology, 42,* 877-888.

Clark, F.J., Larwood, K.J., Davis, M.E., & Deffenbacher, K.A. (1995). A metric for assessing acuity in positioning joints and limbs. *Experimental Brain Research, 107*, 73-79.

Fechner, G.T. (1966). Elements of psychophysics. In D. H. Howes & E. Boring (Eds.), *Elements of psychophysics* (trans. H. E.Adler). New York: Holt, Rinehart, and Winston.

Fernberger, S.W. (1913). On the relation of the methods of just perceptible difference and constant stimuli. *Psychological Monograph, 14*, 1-81.

John, K.T., Goodwin, A.W., & Darian-Smith, I. (1989). Tactile discrimination of thickness. *Experimental Brain Research, 78*, 62-68.

Lee, H., Nicholson, L., Adams, R., & Bae, S-S. (2005). Proprioception and rotation range sensitization associated with subclinical neck pain. *Spine, 30*, E60-E67.

Magill, R., & Parks, P. (1983). The psychophysics of kinesthesis for positioning responses: The physical stimulus-psychological response relationship. *Research Quarterly for Exercise and Sport,* 54, 346-351.

McNicol, D. (2004). *A primer of signal detection theory*. Mahwah, NJ: Erlbaum.

Peirce, C., & Jastrow, J. (1885). On small differences of sensation. *Proceedings of National Academy of Science, 3*, 75-83.

Refshauge, K.M., Kilbreath, S.L., & Raymond, J. (2000). The effect of recurrent ankle inversion sprain and taping on proprioception at the ankle. *Medicine and Science of Sports and Exercise, 32*, 10-15.

Shannon, C.E., & Weaver, W. (1964). *The mathematical theory of communication*. Urbana: University of Illinois Press.

Snodgrass, J.G., Berger, G.L., & Haydon, M. (1985). *Human experimental psychology.* USA: Oxford University Press.

Urban, F. M. (1910). The method of constant stimuli and its generalizations. *The Psychological Review, 17*, 229-259.

Waddington, G., & Adams, R. (1999). Discrimination of active plantarflexion and inversion movements after injury. *Australian Journal of Physiotherapy,* 45, 7-13.

Welford, A. T. (1976). *Skilled performance: Perceptual and motor skills.* Glenview, IL: Scott Foresman.

Willems, T. M., Witvrouw, E., Delbaere, K., Philippaerts, R., De Bourdeaudhuij, I., & De Clercq, D. (2005). Intrinsic risk factors for inversion ankle sprains in females: A prospective study. *Scandinavian Journal of Medicine and Science in Sports, 15*, 336-345.

APPLYING SDT TO A SPEEDED PAPER AND PENCIL TEST

John S. Monahan and Maureen A. Harke
Department of Psychology, Central Michigan University, Mt. Pleasant, MI 48859
john.monahan@cmich.edu

Abstract

Forty-eight women and 48 men completed the Revised Mental Rotations Test (Peters,1995) (RMRT) twice, three weeks apart. Between tests half of them completed six practice tests made from figures similar to but different from those on the RMRT. RMRT has 24 items; each has a target figure and four alternative figures, two rotations of the target plus mirror images or structurally different figures. Items are scored correct if both rotations are marked, or otherwise, incorrect. Standard RMRT scores, sensitivity (d'), and bias (β) were analyzed. It appears that practice aids women more than men and does so by improving rotated pattern recognition and by speeding up the recognition process for some types of figures thus producing higher d' on some types of items and reducing β on all types.

Measured gender differences in cognitive abilities appear to have diminished over time (Wraga, Duncan, Jacobs, Helt, & Church, 2006), but mental rotation, continues to yield large statistically significant differences favoring men. Linn and Peterson (1985) found that mental rotation, continued to yield a large effect size favoring men. A more recent meta-analysis (Voyer, Voyer, & Bryden, 1995) found a linear increase in effect size with increasing age across categories suggesting the possibility that gender differences in spatial ability may be affected by experience, sexual differentiation, or both. The largest effect was found with the Mental Rotations Test (MRT) (Vandenberg & Kuse, 1978).

Sexual differentiation alone fails to take into account how experience influences ability and brain function (Voyer, 1995). In addition to the biological differences between women and men, there are experiential differences between genders (Terlecki & Newcombe, 2005), such as early childhood toy preference and play (Voyer et al., 2000). For example, puzzles are completed by mentally or physically rotating pieces to make comparisons. Dolls and board games do not emphasize spatial relationships and may be less likely to influence spatial ability. Voyer et al. found that men and women who reported childhood preference for spatial toys performed better on the MRT than those who preferred non-spatial toys. Never-the-less, men's performance was greater overall.

If long-term involvement in spatially oriented activities affects spatial ability, short-term experiences may also. The effects of practice are debated among researchers. One idea is that participants retrieve stored stimuli from memory and therefore answer more questions faster and more accurately on repeated tests without improvement on novel stimuli (Heil, Rosler, Link, & Bajric, 1998; Tarr & Pinker, 1989; Wiedenbauer, Schmid, & Jansen-Osman, 2007). Feng, Spence, and Pratt (2007) found that 10 hours practice on an action video game reduced gender difference in the MRT. Another is that practice sessions as brief as one hour may improve participants' performance independent of the orientation of the practiced stimuli indicating an improvement in spatial ability (Murray, Jolicoeur, McMullen, & Ingleton, 1993).

For the MRT and its successor, Peters's (1995) Revised MRT (RMRT), standard scoring procedure is to count as correct each item for which both correct alternatives were marked. This scoring method ignores some of the additional information available: items for which only one correct alternative is marked and items for which foils are marked incorrectly. One way to assess that additional information is via signal detection theory (SDT) analysis.

Method

Participants

There were 96 college students recruited from introductory psychology classes: 48 women and 48 men, 24 of each in the practice and control groups. The practice group took the RMRT, practiced for 6 sessions over three weeks using similar, but different items, and then took the RMRT again. The control group took the RMRT, did not practice and after three weeks took the RMRT again. Practice group participants received course credit plus $25. Control groups participants received course credit plus $5. Practice and control group participants were tested in separate semesters.

Materials

RMRT. The RMRT consists of 24 items, each with a target figure and four alternatives. Two alternatives are rotations of the target; the other two alternatives are foils. Voyer and Hou (2006) classify items as occluded if one of the figures, target or foil, has a section not visible because of a rotation of the figure, or non-occluded if each figure has all sections visible. Eight non-occluded items have two figures, structurally different from the target, as foils; two structurally different foil items have occluded figures. Another eight non-occluded items have rotated mirror images of the target as foils; four mirror image foil items have occluded figures. Two items have mixed foils; one has non-occluded figures; the other has an occluded figure.

Practice Items. Three sets of 24 practice items, five figures per item, were devised using a CAD program. Each figure consisted of 10 cubic sections with at least three right angles projected onto a two dimensional surface. None of the practice figures were identical to any of the figures in the RMRT. Items consisted of a target figure and four alternative figures: two were images of the target rotated on different axes, and two were distracter alternatives: a mirror image of the target and a structurally different figure. Each set of practice items was reordered by randomization, and a reordered set of figures was devised from each item. The original target figure became a rotation of the new target figure, and the order of alternatives was randomized.

Procedure

Participants were administered the initial RMRT. Practice participants were administered two practice sets per week for three weeks beginning the following week. The practice sets for each week consisted of an original set of 24 items and a reordered set of those items with figures within items, including the target, reordered. These were administered on separate days. At the end of each practice session, participants were shown a list of the correct answers to each item. Control group participants did not receive any practice. After the three week period, participants were again administered the RMRT. Correct answers to the RMRT were never shown.

Administration of the tests and practice sets was conducted using the instructions for the RMRT (Peters, 1995), including allowing 3 min for each of the two 12 item sections with a 2 min break between sections. Participants were subsequently debriefed and paid for their service.

Results

Figure 1 shows items correct, d', and β for tests 1 and 2. Tests were scored as the number of items correct, i.e. the number of items in which both correct alternatives were marked, yielding a maximum score of 24. SDT measures were scored using each of the alternatives and foils of each item, thus there were 48 signal present items and 48 signal absent items.

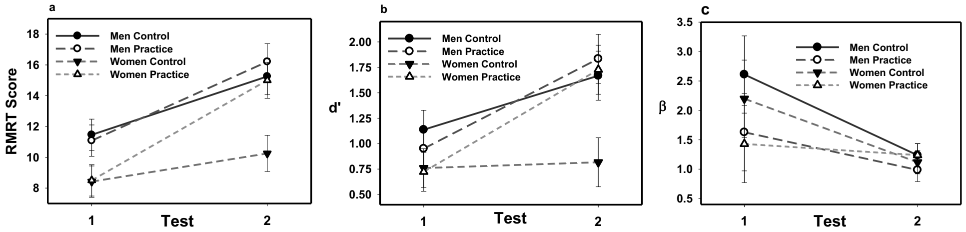


Figure1: Overall results for tests 1 and 2 as a) items correct, b) d', and c) β. Error bars represent SEM.

Similar scoring occurred for each of the three item types, structurally different, mirror image, and occluded alternative, except that proportion rather than number of correct items was used. Initial analyses were conducted using partially repeated measures ANOVA. Condition means were compared using α=.005.

Items Correct. Test, *F*(1,92) = 113.75, p<.0001, sex, *F*(1,92) = 8.93, p=.002, and the interaction between test and practice, *F*(1,92) = 4.24, *p* =.042, had significant effects on items correct. Practice group men had higher scores than women on test 1, but not test 2. Control group men had higher scores than women on both tests. Both women and men attempted about four more items on the second than the first test.

d'. Test, *F*(1,92) = 47.86, *p* < .001, and the interaction between test and practice, *F*(1,92) = 13.17, *p* < .001, had significant effects on *d'*. Sex had a marginal effect, *p*=.052. Practice group men's mean *d'* was not significantly greater than women's on test 1 or 2. Control group men's *d'* was greater than control group women's on test 2. Practice group men and women improved *d'* from test 1 to test 2.

Table 1. *Proportion Correct by Item Type, Test, and Gender for Control and Practice Group Participants*

	Control Group			Practice Group		
Test	Men	Women	d	Men	Women	d
			Structurally Different Foils, Eight Items			
1st	0.54 (0.29)*	0.43 (0.23)	0.44	0.54 (0.26)*	0.36 (0.17)	0.77
2nd	0.70 (0.25)*	0.59 (0.22)	0.45	0.78 (0.21)*	0.69 (0.21)**	0.42
			Mirror Image Foils, Eight Items			
1st	0.54 (0.29)*	0.38 (0.28)	0.58	0.49 (0.29)	0.43 (0.26)	0.23
2nd	0.68 (0.31)*	0.40 (0.32)	0.89	0.66 (0.33)	0.67 (0.31)**	0.03
			Occluded Alternatives, Seven Items			
1st	0.36 (0.19)	0.28 (0.23)	0.37	0.37 (0.25)	0.29 (0.18)	0.35
2nd	0.54 (0.30)*	0.28 (0.26)	0.91	0.60 (0.28)	0.54 (0.26)**	0.25

Note: Standard deviations in parenthesis. d is Cohen's d and reflects gender difference effect size for participants in the same group for each test.

*Men's proportion correct greater than women's from the same group, *p* < .005.

**Practice group proportion correct greater than control group for the same gender, *p* < .005

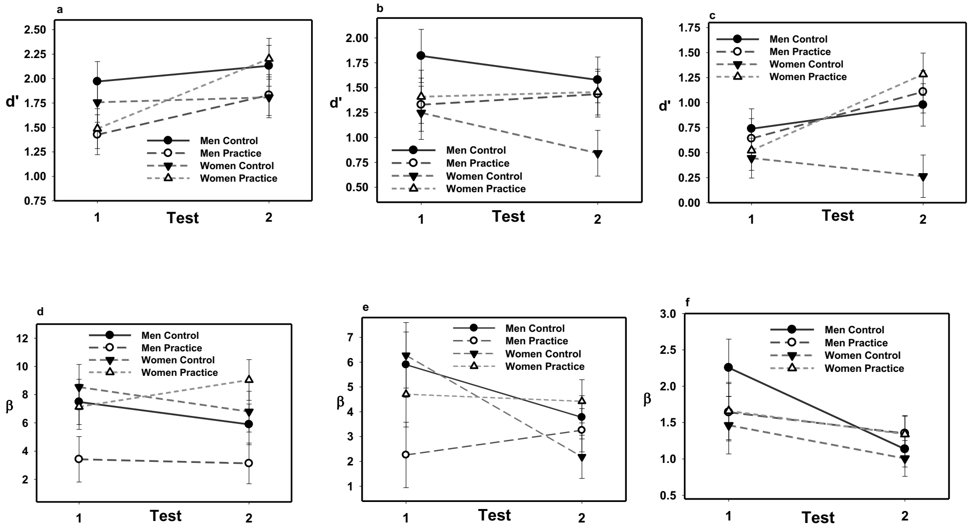


Figure 2: a), b), and c) are *d'* and d), e), and f) are β for structurally different foils, mirror image foils, and occluded alternative items for practice and control group women and men.

β. Only test had a significant effect on β, *F*(1,92) = 8.25, *p* =.005. Overall β on test 1 was greater than on test 2. Thus except for the control group women, Items correct and d' improved from test 1 to test 2, and for everyone, criterion became stricter.

Proportion correct on the three subtypes of items are shown in Table 1. Figure 2 shows *d'* and β for items with structurally different foils, mirror image foils, and occluded alternatives.

Structurally Different Foils. Proportion correct on non-occluded, structurally different foil items were affected by test, *F*(1,92) = 92.85, p< .0001, sex, *F*(1,92) = 8.38, p=.005, and the interaction of test and practice, *F*(1,92) = 6.85, p=.01. With structurally different foil items, although women and men in both groups improved their proportion correct from test 1 to test 2, men scored higher than women in the same group on both tests. The result was different with *d'*; only test had a significant effect on *d'*, *F*(1,92) = 6.68, *p* = .011. Practice group women and men improved their *d'* from test 1 to test 2, and women's d' was greater than men's on test 2. On the other hand only sex had a significant effect on β, *F*(1,92) = 5.97*, p* = .016. Practice group men had a stricter criterion than practice group women on both tests.

Mirror Image Foils. Proportion correct on non-occluded, mirror image foils was affected by test, *F*(1,92) = 31.6, *p*<.0001, sex, *F*(1,92) = 4.94, *p* =.029, and the interaction between test and practice, *F*(1,92) =5.95, *p* =.017. Practice group women and men and control group men improved proportion correct from test 1 to test 2, and control group men out scored control group women on both tests. None of the three factors or their interactions had an effect on *d'*, nevertheless, control group men had higher *d'* than control group women on both tests. Test, *F*(1,92) = 4.39, *p* =.039, and the interaction between test and practice, *F*(1,92) = 6.95, *p* =.01, had significant effects on β. Control women reduced β from test 1 to test 2. Practice group men had a stricter criterion than women on test 1, but control group women had a stricter criterion than men on test 2.

Occluded Alternatives. Proportion correct on occluded alternative items was affected by test, *F*(1,92) = 49.32, *p*<.0001, sex, *F*(1,92) = 7.03, *p* =.009, the interaction between test and practice, *F*(1,92) = 10.19, *p* =.002, and the interaction among test, sex, and practice, *F*(1,92) = 4.17, *p* =.044. Practice group women and men and control group men increased their proportion correct from test 1 to test 2, and control group men scored higher than women on test 2. Both test, *F*(1,92) = 8.82, *p*<.004, and the interaction of test and practice, *F*(1,92) = 7.31, *p*<.008, had significant effects on *d'*. Practice group women improved their *d'* from test 1 to test 2, and control group men had a greater *d'* than did women. Only test had a significant effect on β, *F*(1,92) = 6.72, *p*<.011. Control group women's and men's criteria became stricter from test 1 to test 2.

Discussion

Figure 1 shows that with adequate practice, women can score well on the RMRT, but that previous test taking without practice has little effect on women's performance. This result is confirmed by *d'* data, again women with practice improved their sensitivity to a level similar to men's, but women without practice did not. Interestingly, everyone's criterion became stricter by test 2, regardless of sex or practice. Possibly, all had a better idea of what to look for. Thus, although *d'* includes more of the available data in its calculations, the overall picture it presents is little different from the usual number of correct items.

For specific item types, the results were somewhat different. For structurally different foil items, proportion correct results were different from the overall results: men in both groups scored higher on both tests. These were the easiest items according to proportion correct scores. The *d'* scores for these items showed a different pattern: practice group women and men improvement in *d'* from test 1 to test 2, and practice group women had higher *d'* than men on test 2. Practice group men had a stricter criterion than women on both tests.

For mirror image foil items, proportion correct was higher for control, but not practice, group men than women on both tests. Practice group men and women and control group men improved their proportion correct from test 1 to test 2. Control group women did not. The *d'* scores did not increase for men or women in either group. There was a reduction in β from test 1 to test 2 for control group women. Practice group men had a lower β than women on test 1, but control group women had a lower β than men on test 2.

For occluded alternative items, all but control group women increased their proportion correct from test 1 to test 2. Practice group men scored higher than women on test 2. Practice group men scored higher than women on test 2. Practice group women increased their *d'*, and control group men had higher *d'* then women on test 2. Control group women and men made their criterion stricter from test 1 to test 2.

Both men and women in the practice group increased their overall *d'*, indicating that practice improves overall sensitivity rotated figures. However, for two of the item types, structurally different foils and occluded alternatives, almost all improved d'. Structurally different and mirror image foil items were in the practice items, thus the improvement in practice group women's *d'* on structurally different foil items may be due to practice. On the other hand, there were no occluded figures on the items in the practice sets. Thus, the fact that women improved their *d'* on occluded alternative items cannot be straightforwardly attributed to practice, because they did not practice with such items. If practice was at work, it must have been a general effect, rather than a specific one. Feng et al. (2006) suggest that improvement in women's MRT performance in their experiment was caused by improved spatial attention. Possibly, practice with MRT-like items improves spatial attention, as well as giving practice on some types of items. That no one improved *d'* on mirror image foil items was unexpected because such items were included in practice. An explanation for that phenomenon is not obvious at this time.

Because *d'* measures sensitivity in the absence of any control for speed, it changes a speed test score into a power test score. Thus caution is needed when interpreting the results.

Practice group women scored overall as well as practice group men. Spaced practice of six 15 min sessions appears to be enough to reduce the gender difference on MRT to non-significance on both items correct as well as sensitivity. It is not known, however, whether this gender equality can be sustained without continuing practice.

References

Feng, J., Spence, I. & Pratt, J. (2007). Playing an Action Video Game Reduces Gender Differences in Spatial Cognition. *Psychological Science, 18* (10), 850-855.

Heil, M., Rosler, F., Link, M., & Bajric, J. (1998). What is improved if a mental rotations task is repeated - the efficiency of memory access, or the speed of a transformation routine? *Psychological Research*, *61,* 99-106.

Linn, M.C., & Peterson, A.C. (1985). Emergence and characterization of gender differences in spatial abilities: A meta-analysis. *Child Development, 56,* 1479-1498.

Murray, J.E., Jolicoeur, P., McMullen, P.A., & Ingleton, M. (1993). Orientation-invariant transfer of training in the identification of rotated natural objects. *Memory & Cognition, 21,* 604-610.

Peters, M. (1995). *Revised Vandenberg & Kuse Mental Rotations Tests: Forms MRT-A to MRT-D.* Guelph (ON), Canada: Technical Report, Department of Psychology,

Tarr, M.J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology, 21,* 233-282.

Terlecki, M.S., & Newcombe, N.S. (2005). How important is the digital divide? The relation of computer and videogame usage to gender differences in mental rotation ability. *Sex Roles: A Journal of Research, 433,* 433-441.

Vandenberg, S.G., & Kuse, A.R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47,* 599-604.

Voyer, D. (1995). Effect of practice on laterality in a mental rotation task. *Brain and Cognition, 29,* 326-335.

Voyer, D., & Hou, J. (2006). Type of items and the magnitude of gender differences on the mental rotations test. *Canadian Journal of Experimental Psychology, 60,* 91-100.

Voyer, D., Nolan, C., & Voyer, S. (2000). The relation between experience and spatial performance in men and women. *Sex Roles, 43,* 891-915.

Voyer, D., Voyer, S., & Bryden, M.P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117,* 250-270.

Wiedenbauer, G., Schmid, J., & Jansen-Osmann, P. (2007). Manual training of mental rotation. *European Journal of Cognitive Psychology, 19,* 17-36.

Wraga, M., Duncan, L., Jacobs, E.C., Helt, M., & Church, J. (2006). Stereotype susceptibility narrows the gender gap in imagined self-rotation performance. *Psychonomic Bulletin & Review, 13,* 813-819.

TESTING THE LOGISTIC PREDICTION OF THE RELATIVE JUDGMENT THEORY OF PSYCHOLOGICAL DISCRIMINATION

Stephen Link
University of California, San Diego
link@ucsd.edu

ABSTRACT

The theory of comparative judgment proposed by Link (1975, 1992) received various tests and confirmations through the joint analysis of response times and response probabilities. These are strong tests of the general character of such a stochastic process. However this paper creates and applies a new approach to testing the prediction that the form of choice probabilities is logistic.

At the foundation of psychophysical theories of comparative judgment are assumptions regarding the characterizations of stimuli and how these are used to form a judgment. Fechner (1860, see Link 1994) represented stimuli by internal variables postulated to have Gaussian distributions because these distributions captured the 19th century idea about how error disturbed measurements, whether physical or mental.

The idea was sufficiently powerful to lead Fechner into a theory of comparative judgment that created the basis for experimental psychology (Link, 1994) as well as statistical hypothesis testing. The theory supposes that a particular internal stimulus value is compared to an internal threshold or criterion. Whether the internal stimulus value is above or below this threshold determines the choice response made by the subject. However, the theory was not able to account for the relations between response time and response probability, in part because the theory itself has nothing to do with the unfolding of a decision in time. What tests there are showed failure of the assumption that distance from the threshold or criterion determined response time. (e.g., Thomas and Myers, 1968).

The consideration of time-dependent mental processes begins with an entirely different view of the mechanism for creating a choice between two alternatives. In keeping with the general ideas introduced by Abraham Wald (1947) in the sequential analysis of statistical hypotheses Link introduced a distribution-free sequential theory of comparative judgment (Link and Heath, 1975). The theory's strength was in predicting relations between response time and response probability. Many tests showed the theory to provide an accurate portrayal of how subjects made judgments in many choice experiments.

Link (1978) showed that the response probabilities predicted by the theory had the form of a logistic function. This particular function was used by statisticians to fit data. But the source of the logistic equation was unknown. Its close fit to results in many different scientific areas suggested that there must be a common basis for its frequent appearance. The surprise was that this function describing choice probabilities was an outcome of the random walk theory of statistical hypothesis testing due to Wald (1947) and extended by Link (1978).

This logistic function depends on two parameters and defines the probability of choosing between one of two response alternatives as

$$P = \frac{1}{1+e^{-\theta A}} \tag{1}$$

where θ is a measure of discriminability and A is the accumulated amount of comparative difference between the stimulus and a referent needed to trigger a response. The derivation and proof that this logistic function is a consequence of random walk theory is found in Link (1978).

The parameter θ that captures discriminability is derived from the formal analysis of the underlying stochastic accumulation of comparative differences through an application of Wald's Identity. Although Link (1978) derived the logistic response function as a consequence of bounded random walk theory, the parameter θ remained a discrimination parameter without formal relation to the particular nature of stimuli under judgment. In this sense the derivation did not depend on the form of the underlying probability distributions thought to characterize the stimuli.

A simple calculation shows that the unknown parameters θ and A are jointly determined by computing

$$\ln\left[\ln\left(\frac{P}{1-P}\right)\right] = ln\theta + lnA. \qquad (2)$$

This addition of parameters, even on a logarithmic scale, suggests performing experiments designed to cause independent changes in θ and A in order to test whether the derived logistic form of the choice probabilities results from such experimental manipulations. The analysis of such an experiment is the focus of this paper.

The experiment occurred before these theoretical results were known. Link and Tindall (1971) and Link (1971) required subjects to compare sequentially presented horizontal line segments on a computer controlled display, and to judge whether a comparison line was the same or different from a fixed standard. Another change in performance resulted from requiring the subjects to respond under three different instructions regarding the speed and accuracy of their responses. Subjects were to respond as accurately as possible, to "Beat" a 460 msec response time deadline while being as accurate as possible, and to "Beat" a 260 msec response time deadline while being as accurate as possible.

Four well-practiced subjects made choice judgments in 60-trial blocks within which the size of the comparative difference remained fixed as did the instructions on speed and accuracy. Within each block the comparison stimulus was either the same or different from the 2cm standard on 50 % of the trials. The first ten trials, with 50% same and 50% different comparisons, were treated as practice and do not enter into the analysis below. Each subject used the same speed and accuracy conditions during a day in which eight blocks of trials totaling 8*50=400 test trials yielded 100 trials for each of four different sizes of comparative difference, .1cm, .2cm, .3cm and .4cm. Each speed-accuracy condition ran for four successive days. Thus each subject contributed 1600 judgments for each speed-accuracy condition consisting of 400 judgments for each level of stimulus difference. The total number of test trials is 19,200.

The judgments proved to be quite similar in the probabilities of a correct response regardless of whether the response was "Same" or "Different" from the standard 2cm line (cf, Link, 1992, pp 214-223). For this reason, and to keep with the results as reported by Link and Tindall (1971), the response probabilities presented below are for the probabilities of a correct response whether the judgment was "Same" or "Different." Each row of Table 1 corresponds to a speed-accuracy condition for which there are a total of 6400 observations. Each column corresponds to a fixed amount of stimulus difference and a total per column of 4800 observations. As might be expected various amounts of stimulus difference and the speed-accuracy conditions caused large changes in the response probabilities.

Table 1: Observed and Predicted Correct Response Proportions

	1cm		2cm		3cm		4cm	
Speed-Accuracy	Obs.	Pred.	Obs.	Pred.	Obs.	Pred.	Obs.	Pred.
ACCURACY	.804	.834	.919	.918	.971	.965	.988	.983
460 msec	.731	.734	.818	.807	.876	.877	.905	.916
SPEED	.609	.595	.630	.637	.676	.684	.714	.720

The question about the adequacy of the logistic equation to describe these performances is best answered by analysis of individual subject data, rather than an analysis of the average data provided in Table 1. The same calculations can be applied in either case. To illustrate the ideas the probabilities in Table 1 are used below. Then the method will be applied to all four subject's individual results.

The additivity of effects due to θ and A requires a new form of analysis. By computing $\ln[\ln(P/(1-P))]=\ln\theta + \ln(A)$ all response proportions within each cell are a consequence of a sum of values dependent upon θ and A. Presumably these parameters depend upon the amount of stimulus difference, θ, and the amount of comparative difference, A, required to create a response. Thus, for each of J columns there is a value of θ_j. However, noting that these values may be averaged across a row i, (i=1,…,I) yields an average for row i of,

$$\frac{1}{J}\sum_{j=1}^{J}[\ln(\theta_j) + \ln(A_i)] = \ln(A_i) + \theta^* \tag{3}$$

where $\theta^* = \ln(\theta_1\theta_2 \ldots \theta_J)^{1/J}$ the logarithm of the geometric mean of the unknown θ values. Similarly, averages down a column produce,

$$\frac{1}{I}\sum_{i=1}^{I}[\ln(\theta_j) + \ln(A_i)] = \ln(\theta_j) + A^* \tag{4}$$

where $A^* = \ln(A_1A_2 \ldots A_I)^{1/I}$ the logarithm of the geometric mean of the unknown values of A. The overall mean, M, equals θ* + A*. These ideas are shown in Table 2. Here are the various probabilities in Table 1 converted to ln(ln()) values with averages shown in the penultimate right-hand column and the next to the lowest row. The overall mean M = 0.335.

Table 2 ln(ln(P/(1-P)))

	1cm	2cm	3cm	4cm	Average	Estimate of
ACCURACY	0.345	0.887	1.256	1.484	0.993	$\theta^* + \ln A_{ACC}$
460	0.000	0.407	0.670	0.813	0.473	$\theta^* + \ln A_{460}$
260	-0.814	-0.631	-0.307	-0.089	-0.460	$\theta^* + \ln A_{260}$
Average	-0.157	0.221	0.540	0.736	0.335	$=M=\theta^*+A^*$
Estimate of	$\ln\theta_1 + A^*$	$\ln\theta_2 + A^*$	$\ln\theta_3 + A^*$	$\ln\theta_4 + A^*$		

Although all the parameter values are unknown, a test of the logistic equation is still possible. Note that the sum of averages for a particular row and column (i, j) gives

$$\begin{aligned}\theta^* + \ln A_i + A^* + \ln\theta_j &= \ln A_i + \ln\theta_j + \theta^* + A^* \\ &= \ln A_i + \ln\theta_j + M.\end{aligned} \tag{5}$$

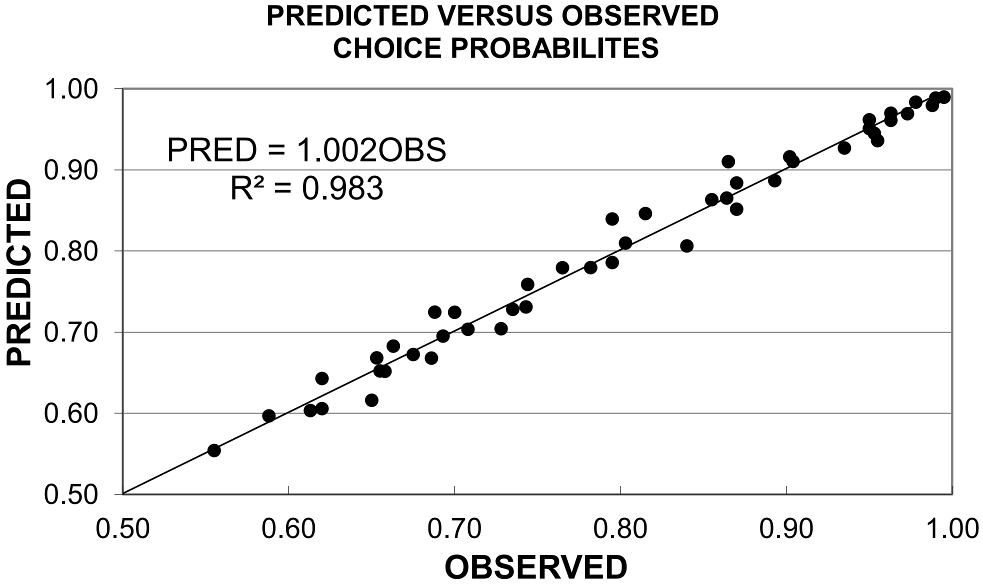


Figure 1. Response probabilities observed and predicted for the experiment of Link and Tindall (1971) employing speed-accuracy instructions and changes in discriminability.

Thus, by subtracting M from the sum of a row and column averages and exponentiating the result, the unknown value of $\theta_j A_i$ is estimated. These may then be used in Equation 1 to predict the response probabilities. If the logistic model is correct then there should be a close correspondence between the predicted and observed probabilities. Notice too that this is not trivial, the same value for θ is assumed to apply across the speed-accuracy conditions and a fixed value for A is assumed to apply across values of stimulus difference.

For the purpose of illustration Table 1 includes predicted values for these marginal response probabilities, and the fit is quite good. However, a better test is to apply the method to each subject separately and then examine the overall comparison between observed and predicted values. These values for the 12 probabilities for each of the four subjects appear in Figure 1. The linear equation of best fit with fixed zero intercept shows good agreement between observed and predicted response probabilities. The average deviation from predicted values is 0.002.

The good agreement between predicted and observed probabilities suggests carrying this analysis a step further. Notice that each row or column average may have the value M subtracted from it to leave only those parameters related to the experimental condition assigned to either the row or column. That is, for column 1:

$$\text{Average Column } 1 - M = A^* + \ln\theta_1 - (\theta^* + A^*) = \ln\theta_1 - \theta^*. \qquad (6)$$

This value equals the logarithm of θ_1 minus the logarithm of the geometric mean of the θs. Exponentiating this result yields the value of θ divided by the geometric mean of the theta values – a relative measure for θ.

These values for the average results and for averages of subject's relative parameters are shown in Figure 2. These are relative values of the discrimination parameter θ for each level of stimulus difference, .1cm, .2cm, .3cm and .4cm. Of course these values increase as the size of the stimulus difference increases.

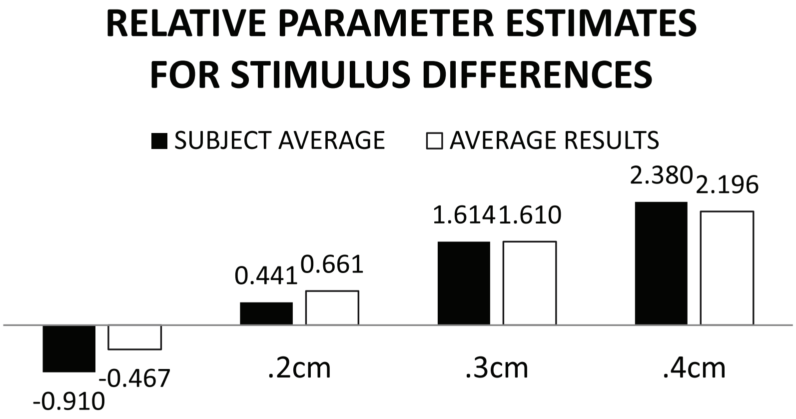


Figure 2. Relative estimates of the effects of stimulus difference on performance. The first column corresponds to .1cm. Averages of subject estimates agree closely with those obtained from the average results in Table 1.

The real question is how these values are related to the physical values of the stimuli being compared. The fact that this ratio is a dimensionless quantity suggests that the numerator and denominator be estimated by using the actual stimulus values.

There are no physical values for the parameter A, thought to change under changes in the speed-accuracy instructions given the subjects. The physical parameters corresponding to the instructions may be related to a subject's perception of how much stimulus difference must be accrued to make a response or in some way to the time to respond. But at this time the association with the relative speed-accuracy parameters shown in Figure 3 is unknown. The idea that the amount of accumulated stimulus difference needed to respond is the measure of the parameter A suggests that A should increase as the speed deadlines are relaxed, and this is quite clear in Figure 3.

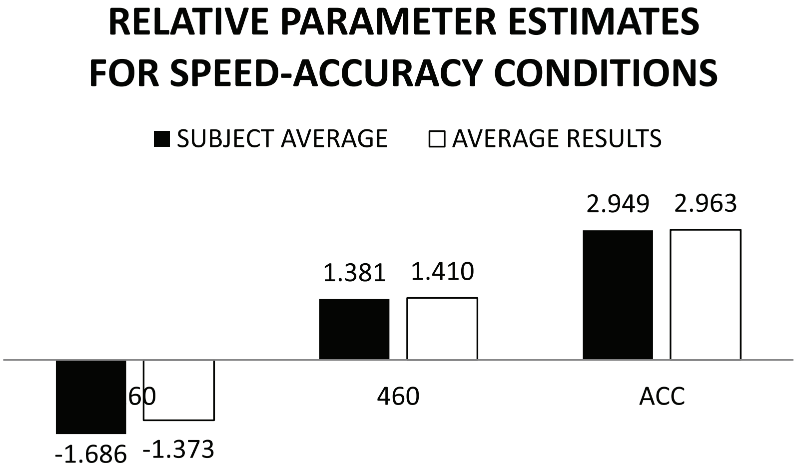


Figure 3. Relative values of the parameter A controlling the average duration of responses.

This next step toward understanding the relationship between the theoretical parameters and the physical values that underlie them requires a deeper analysis of the underlying probability distributions representing the stimuli and the particular mechanism for creating a comparative difference. The small amount of space available here does not allow for this extended discussion.

However, these tests of the predicted logistic representation of response proportions seems sufficiently strong to provide excellent support for the sequential theory of psychological discrimination first proposed in Link and Heath (1975).

References

Fechner, G. T. (1966). *Elements of psychophysics. Vol 1.* (E. G. Boring and D. H. Howes, Eds. H. E. Alder, Trans.) New York; Holt Rinehart & Winston. (Original work published 1860).

Link, S. W. (1971). Applying RT deadlines to discrimination reaction time. *Psychonomic Science,* 25 (6), 355-358.

Link, S. W. (1975). The relative judgment theory of two choice response time. *Journal of Mathematical Psychology*, 12, 114-135.

Link, S. W. (1978). The relative judgment theory of the psychometric function. In J. Requin (Ed.) *Attention and Performance VII.* (pp. 619-630). Hillsdale, NJ. Lawrence Erlbaum Associates.

Link, S. W. (1992). *The wave theory of difference and similarity.* Hillsdale, NJ. Lawrence Erlbaum Associates, Publishers.

Link, S. W. (1994) Rediscovering the past: Gustav Fechner and Signal Detection Theory. *Psychological Science*, 5,6), 335-340.

Link, S. W. and Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*. 40, 77-105.

Link, S. W. and Tindall, A. B. (1971). Speed and accuracy in comparative judgments of line length. *Perception and Psychophysics*, 9, 3A, 284-288.

Thomas, E. A. C. and Meyers, J. L. (1971). Implications of latency data for threshold and non-threshold models of signal detection. *Journal of Mathematical Psychology*, 9, 253-285.

Wald, A. (1947*). Sequential analysis.* New York: Wiley.

DOES PSYCHOLOGY REQUIRE MORE (REFINED) DATA?

Denis Cousineau
School of psychology, University of Ottawa, Ottawa, Canada
Denis.Cousineau@UOttawa.ca

Abstract

Experimental psychology is a discipline that collects data to advance knowledge, foster theories and falsify wrong leads. It is a very fertile scientific field, collecting possibly billions of new data every year. Researchers often struggle hard to come up with innovative designs but this richness in methodological sophistication is not mirrored in the behavioral measures collected. Too often, research articles report preferences (or percent correct if there is a best choice) and mean response times. How many experiments, as clever as they might be, will it take to unravel the mysteries of the mind when there are only two variables observed? Here, we will argue that a lot of information is wasted in psychology. Even if it all starts with preferences and response times, averaging is only the beginning. We will discuss the importance of variances and skew in response times. We will also argue in favor of coefficients of changes. These are just examples; we wish to encourage researcher to use the same amount of imagination on measurement as on experimental design to stop the big datawaste.

Psychology and psychophysics are not easy fields of research. Researchers in those fields aim at finding the rules of perception and decision making. Yet, a percept is not a clearly defined phenomenon. Decision-making, presumably based on evidence gathering, is also elusive. Learning and training effects are likewise hard to define: do they imply local or global changes. Neurons implement these abilities. Yet, we cannot define what the *carrier of information* is across cells: is it the spikes, the synchronous firing, the strenght of activation or the time of activation? We are not even sure that neureons alone are responsible for our cognition, some suggesting that astrocytes may play an important role in learning.

Within such a loosely defined framework, it is important to support a multiplicity of approach, both from a methodological perspective and from a quantitative perspective. Yet, whereas we devote much energy to devise wise experimental manipulation, we cannot say the same regarding quantitative analyses. A lot of research papers only examine percent correct and mean response times. How many experiments will it take to uncover the rules of the mind when we use only two measures? It is time that we delve deeper into the data.

Considering that experimental psychologists collect in the order of billions of data per year, it represents a huge investment. We can approximate the cost of one datum in psychophysics and cognitive psychology at about 1 cent (counting the apparatus and the wages), which is affordable, but in total, it represents tens of millions of dollars spent gathering information.

Here, we will argue that we must invest the same ingenuity deriving meaningful measures. We do so by highlighting a few examples.

Beyond the mean

An intriguing paper was published a decade ago "What can a million trials tell us about visual search" (Wolfe, 1998). With so many data, the expectations were high that we would finally know whether visual search for a target proceeds in serial or in parallel. Sadly, all the response times (RTs) were aggregated using the mean so that in effect, there was much fewer data than announced. And the conclusion turned out to be a disappointment: both serial and parallel processing were possible considering the range of means observed. The only true conclusion of that study must in reality be that the means are not informative enough regarding this issue. More data do not represent more information when irrelevant results are kept.

One reason is that the standard Serial Self-Terminating Search model (SSTS), one of the two class of model tested, is based on two unknown parameters, but only one prediction was examined, a prediction based on the mean (see below). Hence, the data under determined the model.

The prediction tested is the ratio of increase in means for target absent trials as display size increases (*D*) relative to the increase in means for target present trials. Let $\mathbf{RT} = \sum_{i=1}^{d} \mathbf{T}_1 + \mathbf{T}_0$ where *d* is the number of locations visited (variable from trials to trials, but depending on the total number of locations *D*), $\mathbf{T}_1$ is the time to scan and decide target presence or not at one location, and $\mathbf{T}_0$ is the residual time (motor and perceptual processing). Both $\mathbf{T}_0$ and $\mathbf{T}_1$ are unknown parameters. Hence, $E(\mathbf{RT}|D) = E(\mathbf{d})E(\mathbf{T}_1) + \mathbf{T}_0$. The change in mean from a high display size condition relative to a low display size condition is

$$\Delta E(\mathbf{RT}|D \text{ vs. } 1) = E(\mathbf{RT}|D) - E(\mathbf{RT}|1) = (E(\mathbf{d}) - 1) \times E(\mathbf{T}_1) \tag{1}$$

Contrasting target present (self-terminating processing for which $E(\mathbf{d}) = (D+1)/2$) and target absent conditions (exhaustive processing for which $E(\mathbf{d}) = D$), we get

$$\frac{\Delta E(RT|D \text{ vs. } 1, \text{ target present})}{\Delta E(RT|D \text{ vs. } 1, \text{ target present})} = \frac{(D-1)E(\mathbf{T}_1)}{(\frac{D+1}{2} - 1)E(\mathbf{T}_1)} = 2 \tag{2}$$

This is the famous 2 to 1 ratio of target absent to target present slopes. It is a pointwise prediction regarding the ratio of changes in means. More importantly, this measure based on mean response times is independent of the two unknown parameters $\mathbf{T}_0$ and $\mathbf{T}_1$.

As seen from this example, means can be assembled to create *second-order* measures. Assuming a specific model, here the SSTS, this measure, the ratio of changes in mean, is predicted to be a constant and this is an easy-to-test prediction. It is based on changes because it is the best solution to remove residual processes' influences.

Other descriptive statistics can be used instead of the mean. For example, the change in variance for target absent trials is predicted to be a linear function of *D* according to the SSTS model.

Likewise, skew is expected to tend towards zero as display size is increased. This is predicted under SSTS by the Central Limit Theorem: As the number of location increases, distribution of RT should tend toward a normal distribution.

Finally, minimum RT can be used to make predictions. Indeed, for many parallel models, mean and variance can increase, but the best RT should be constant. Hence, the first percentile (if the number of observations permits) should be constant across conditions under such a model.

These last sources of information are rarely, if ever, mentioned in the Result sections of published research papers. Further, this information cannot be deduced from inspection of the mean results (or inspection of the means and variances if both aspects of the data are provided). Hence, a potentially important source of information is annihilated in the publication process.

Coefficient of change in variations

Means and variances can also be mixed in what are called Coefficient of change in variation relative to change in means (noted in short ΔCV).

Within SSTS, we get a nice prediction is we suppose that the variance of $\mathbf{T}_1$ is small:

$$\begin{aligned}\Delta CV(\mathbf{RT}) &= \frac{\sqrt{\Delta Var(\mathbf{RT}|D \text{ vs. 1, target present})}}{\Delta E(\mathbf{RT}|D \text{ vs. 1, target present})} \\ &= \frac{\sqrt{Var(\mathbf{d})E^2(\mathbf{T}_1) + Var(\mathbf{T}_0) - Var(1)E^2(\mathbf{T}_1) - Var(\mathbf{T}_0)}}{E(\mathbf{d})E(\mathbf{T}_1) - E(1)E(\mathbf{T}_1)} \\ &\approx d\frac{\sqrt{\frac{D^2}{12}E^2(\mathbf{T}_1)}}{\frac{D+1}{2}E(\mathbf{T}_1)} = \frac{2}{\sqrt{12}}\frac{D}{D+1} \approx 0.577\end{aligned} \tag{3}$$

because $Var(\mathbf{d}) = \frac{(D-1)(D+1)}{12} \approx D^2/12$ and for D large, $D/(D+1)$ tends toward 1. This prediction is again a pointwise prediction independent of the parameters $\mathbf{T}_0$ and $\mathbf{T}_1$. By using subtraction, it removes the residual times and by using a ratio, it removes the time of a single scan-and-decide process.

In a different task, this second-order measure can also make a useful prediction. In the redundant target detection task (Miller, 1982), RT is the time to locate one target attribute, but more than one can be presented. Detecting a single target attribute requires reaching a threshold *k* so that an asymptotic counting model predicts that $\mathbf{RT} = \min_{k:r\times N}(\mathbf{T}_1) + \mathbf{T}_0$ where *r* is the number of redundant attributes, *N* is the total number of channels activated by a target attribute (*N* considered large) and $\mathbf{T}_1$ is the time to register single evidence. Hence,

$$E(\mathbf{RT}|r) = \frac{E(\mathbf{T}_1|1)}{\sqrt[k]{r}} + E(\mathbf{T}_0) \quad \text{and} \quad Var(\mathbf{RT}|r) = \frac{Var(\mathbf{T}_1|1)}{\sqrt[k]{r^2}} + Var(\mathbf{T}_0) \tag{4}$$

so that, assuming again that the variance of the residual processes is small,

$$\Delta CV'(\mathbf{RT}) = \frac{SD(\mathbf{RT}|r) - SD(\mathbf{RT}|1)}{E(\mathbf{RT}|r) - E(\mathbf{RT}|1)} = \frac{SD(\mathbf{T}_1|1)}{E(\mathbf{T}_1|1)} = constant \tag{5}$$

In the redundant target detection task, we have predictions on skew, on minima and on the coefficient of change. However, there is not a single interesting prediction regarding the mean RTs. Mean response times are sometimes overrated.

Distributions

Going beyond the mean is just a first step. Variance (or standard deviations) can be used to make interesting predictions. More importantly, they can be combined into coefficients of change that are convenient to make predictions independent of the parameters. Finally, skew and minima are also important measures to test many parallel and serial models.

To go further, we can also make predictions on whole RT distributions. The Miller inequality (1982) is one such prediction in the redundant target detection task. It turns RT distributions in some baseline conditions into a bound use to test RT distributions in

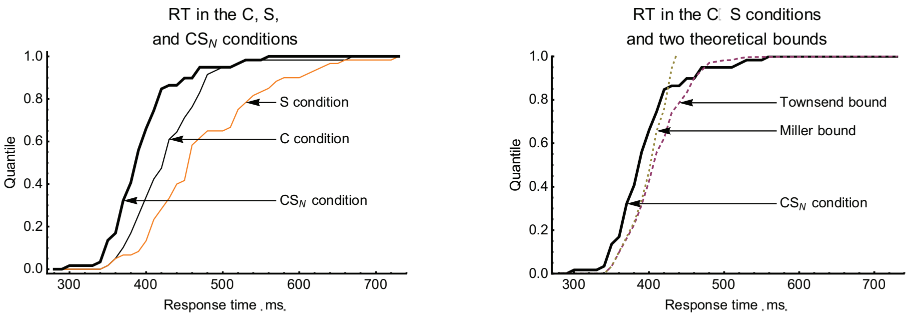


Figure 1. Example data from one participant in Cousineau, Engmann and Mestari (submitted) showing RT distributions for the two baseline conditions (only one target attribute is present, S or C) as well as the RT distribution in the critical CS condition). On the right, we see two boundaries built from the baseline conditions, the Miller bound and the Townsend bound.

conditions with high level of redundancy. This bound is useful to test some models of parallel processing.

Figure 1 shows example data (taken from Cousineau, Engmann and Mestari, submitted) with the Miller bound as well as the Townsend bound (Engmann, 2009). The observed cumulative distribution shows faster RTs than predicted, an indication supporting pooling of activation across the attribute detectors.

Likewise, the Δ plots are nice tools to make simple predictions. This plot shows the difference between the quantiles of two experimental conditions as a function of the mean between the quantiles.

Figure 2 shows three examples of Δ plots in the redundant target detection task. It is based on the triply redundant stimuli vs. the baseline conditions in which only one attribute is presented. See Schwarz and Miller (2012) for more details on this plot.

It is possible to show that in the visual search task, SSTS makes predictions regarding the target absent trials. It predicts that the slope of the Δ plot will be 1. This prediction is based on the technical assumptions that the distribution of processing times is part of the location-scale family of distribution. This assumption is taken from granted by researchers doing vincentizations. However, it is not clearly demonstrated.

For the redundant target detection task, the predicted relation is linear as well, but the slope depends on a free parameter:

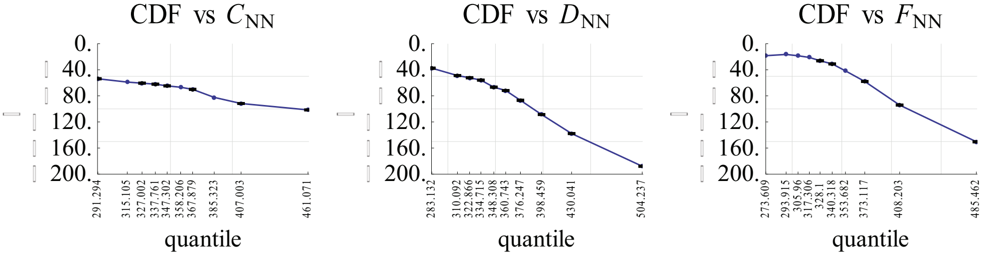


Figure 2. Examples of Δ plots based on the group results of Cousineau, Engmann, & Mestari (in press). The plot shows 9 points, one for each of the quantiles 0.05, 0.15, 0.25, …, 0.95. For each point, we see the difference between the slowest condition and the fastest condition on that quantile as a function of the mean of both quantiles.

$$\frac{\Delta Q(p)}{\Delta E(Q(p))} = 2\frac{\sqrt[\gamma]{r}-1}{\sqrt[\gamma]{r}+1} \tag{6}$$

in which r is the amount of redundancy and γ is the skew in the target detection RTs when there is only one attribute. As an illustration, for the γ offering the best fit to the RTs in the single attributes conditions ($\hat{\gamma} = 1.76$), we find a predicted slope of 0.605 when contrasting baseline conditions with the triple redundant condition and a predicted slope of 0.388 when contrasting baseline conditions with the double redundant conditions. The empirical results are 0.680 and 0.489 respectively.

Again, these are nice predictions that are obtained not just using one or two means, but using a whole range of quantiles. Hence, these are predictions that applies to the RTs, whether they are slow (e.g. the first quantile or below), medium (e.g., about the median) or fast (e.g., above the third quantile). Whereas alternate models could predict similar means, it is difficult to image strikingly different models predicting the same quantile dispersion.

Conclusion

In mechanics, physicists were faced with a similar problem: the two measures they could access were the position (*x*) of a mobile at certain times (*t*). It is exactly analogous to what we have in behavioral psychology where we have only access to response choice and response time. However, early physicists overcame this limitation by deriving measures of a higher order, such as speed and acceleration. These concepts turned out to be fertile and nowadays, no one could imagine mechanics without them.

Psychologist must do the same and find the equivalent of speed and acceleration regarding perception and decision making. When they are found, it will be possible to extract the maximum of information out of the raw data that are decision times and decision choices.

Here we have examined possible candidate measures. Among others, the coefficient of changes may be fruitful as it removes the residual times using a subtraction. At least, it turned out that for many models, it is possible to derive the theoretical value of the coefficient of change. Hence, they synthetize in a compact form some signature of the data that would go unnoticed if only the mean results were reported.

The overall message is simple: stop the data waste. Extract as much information from your results as you can by going beyond the mean results. The definitive measures may not be known at this time, but we should be encouraged to look for useful measures in any possible ways. Quantitative investigation of your results should take as much energy and ingenuity than experimental designs. As a final recommendation, always find a way to make your raw data available to the research community (web sites are less expensive than data), so that future propositions could be benchmarked on existing, relied upon, data sets.

Acknowledgements

This research was supported by grants from the CRSNG and the University d'Ottawa. I would like to thank the members of the Quibb joint laboratory for discussions and commentaries.

References

Cousineau, D. & Larochelle, S. (1997). *PASTIS: A Program for Curve and Distribution Analyses*. Behavior Research Methods, Instruments, & Computers, 29, 542-548.

Cousineau, D., Engmann, S., & Mestari, Z. (2012). *Evidence for parallel coactivation with double and triple redundant targets*. (submitted for publication), submitted, 1-59.

Engmann, S. (2009). Redundancy gain: manifestations, causes, and predictions. Université de Montréal: unpublished doctoral dissertation.

Miller, J. (1982). *Divided attention: Evidence for coactivation with redundant signals*. Cognitive Psychology, 14, 247-279.

Schwarz, W., Miller, J. (2012). *Response time models of delta plots with negative-going slopes*. Psychonomic Bulletin & Review, 19, 555-574.

Wolfe, J. M. (1998). *What can 1 million trials tell us about visual search?*. Psychological Science, 9, 33-39.

THE EFFECT OF NON-ADJACENT LUMINANCE GRADIENTS ON SIMULTANEOUS LIGHTNESS CONTRAST

Daniele Zavagno and Olga Daneyko
Psychology Department, University of Milano-Bicocca, Milano, Italy
daniele.zavagno@unimib.it – olga.daneyko@gmail.com

Abstract

In this work we revisit a phenomenon presented by Agostini & Galmonte (2002), who showed an enhanced effect of simultaneous lightness contrast (SLC) by embedding targets within the glare effect and its photometric negative (Zavagno 1999). In our study we employed 3 manipulations of SLC with targets surrounded by non-adjacent positive luminance ramps (these determine the glare and dark hole effects), negative ramps (the vector of the gradient is inverted 180 deg), and solid black (white background) and white (black background) squares. Configurations with positive ramps show a strong contrast enhancement on both backgrounds with respect to classic SLC (data collected in a similar setup but with different subjects), in line with findings reported by Agostini & Galmonte. The magnitude of the effect is more than halved with negative gradients and solid squares. Findings are consistent with the hypothesis that luminance gradients are relevant information for brightness perception (illumination, luminosity).

The phenomenon dubbed as simultaneous lightness (or brightness) contrast (from now on SLC) is somewhat of a benchmark test for theories that intend to explain how we perceive achromatic colours, i.e. those colours that belong to an ideal grey scale ranging from black to white. In its classic textbook form (Figure 1a), the phenomenon consists of two adjacent squares, one black the other white, which serve as backgrounds (or inducing fields) to two smaller grey squares (from now on targets) that are photometrically equal; in such a configuration, the target seen against the black background looks somewhat lighter than the target seen against the white background.

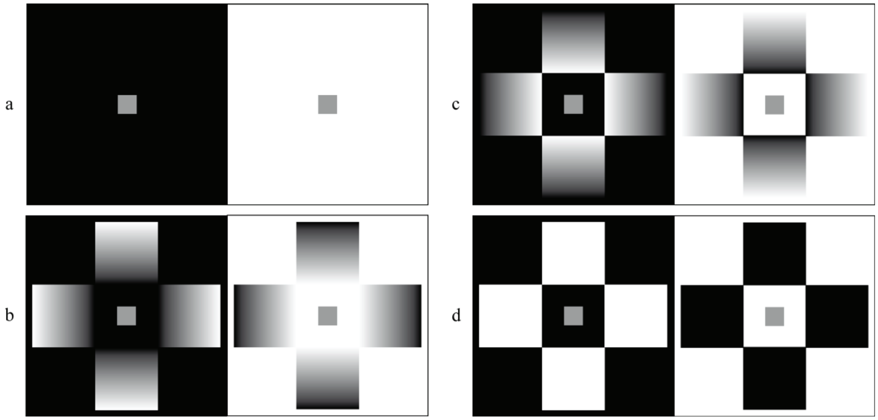


Fig. 1. a) Classic simultaneous lightness contrast; b) SLC with positive ramps; c) SLC with negative ramps; d) SLC with solid squares.

While SLC has been used and studied in many ways over the past 100 years, here we consider a configuration proposed by Agostini and Galmonte (2002), who combined the classic SLC configuration with the glare effect and its photometric negative (Zavagno 1999). They reported that targets equal in luminance surrounded by luminance ramps (as in Figure 1b) appear more different in lightness than similar targets surrounded by solid grey squares. Our experimental study stems from their research.

In particular, we wanted to verify how luminance ramps affected SLC by employing three types of modified SLC displays: with positive ramps (Figure 1b), with negative ramps (Figure 1c), and with flanking solid squares (Figure 1d). We define as positive ramps those linear luminance gradients in which the final luminance value facing a target is equivalent to the luminance of the target's background. Positive ramps determine relevant modifications in the brightness appearance of backgrounds (Zavagno 1999), inducing an impression of luminosity with white backgrounds (i.e. the glare effect), and a darkness enhancement with black backgrounds. Negative ramps are luminance ramps rotated by 180° with respect to positive ramps.

Method

Because many psychology students, i.e. our subject pool, are familiar with classic SLC, it is not rare to find participants reporting no contrast effect when they are presented with the illusion. This is a cognitive bias that goes by the name of "stimulus error" (Kanizsa 1979), according to which a participant to an experiment may want to report what he/she knows or assumes about the physical properties of a stimulus rather than what is actually perceived. In our attempt to avoid, or at least reduce, the occurrence of such a bias, we employed two reflectance values for our targets instead of one as usual. Hence in some configurations targets had the same reflectance, in others they differed in reflectance. This method was already used by Zavagno et al. (2011) and proved to be valid.

Participants

Participants were 35 students (12 male; age range 18-29) from the Psychology Department of the University of Milano-Bicocca. Participants were naïve to the purpose of the experiment.

Material

Stimuli consisted in modified SLC paper displays similar to those depicted in Figure 1b-d. Black and white backgrounds were squares (9.9 deg side); square targets (1 deg side) were inserted at the centre of each background in a square region (1.6 deg side) delimited by four squares (1.6 deg side) that could be either positive ramps, negative ramps, or solid black (on a white background) and solid white (on a black background) (from here on "solid ramps"). The luminance values of the white and the black backgrounds were respectively 2400 and 68.7 cd/m^2. The luminance of positive and negative ramps ranged from 68.7 to 2400 cd/m^2; the luminance of solid white and solid black squares measured respectively 2400 and 68.7 cd/m^2. Targets were cut from Neutral Value Munsell papers 6.0 (reflectance 30%; luminance 900 cd/m^2) and 6.5 (reflectance 36.2%; luminance 1090 cd/m^2). A modified SLC configuration could comprise two targets equal in reflectance (configurations *same*) or two targets different in reflectance (configurations *different*).

Summarizing, variables where: Background (*B*: black, white) × Ramps (*R*: positive, negative, solid) × Target (*T*: 6.0, 6.5) × Target Combination (*TC*: same, different). The combination of all factors determined 24 target-ramp-background combinations, which lead to 12 configurations structurally similar to those depicted in Figure 1b-d.

We employed a matching method with an achromatic Munsell scale, often used in lightness studies. The matching scale consisted in a 16 step Neutral Value Munsell scale ranging from 2.0 to 9.5 in Munsell values, placed on top of a black-white chequered background. Each step of the scale measured 3.01×1.1 deg; the chequered background measured 5.6×19.9 deg, with square checks measuring 0.78 deg per side. Munsell notations were placed under each step on a white stripe 0.4 deg below the row of Munsell steps.

Configurations were viewed one at a time, and were positioned 6 deg above the Munsell scale used for the lightness matching task. Stimuli and the scale were illuminated by the same spotlight (Spotlight Mini Profile ME with 20° objective, 250 W, colour temperature 5900 K), which was the only source of illumination inside the laboratory during the experiment. However, due to the geometry of the spotlight projection, targets and corresponding Munsell values on the scale were different in luminance (scale values: 6.0 = 791 cd/m^2; 6.5 = 980 cd/m^2). The resulting luminance differences between SLC targets and their corresponding Munsell values on the matching scale is, nevertheless, a negligible factor in this experiment because constant with all configurations.

Procedure

With factor *T* as a nuisance variable, we conducted a within subject experiment, while traditionally similar experiments are most often between subjects designs (e.g. Agostini & Galmonte, 2002; Economou, Zdravkovic, & Gilchrist, 2007) to avoid “learning” effects.

The participant entered the laboratory illuminated only by the spotlight, and was seated at a distance of 114 cm. Viewing distance was secured by a chinrest. The participant was informed that the task consisted in finding the grey match of each square target in a configuration. If a perfect match could not be found for a target, the participant was instructed to indicate the Munsell step that was closest to its grey appearance. The modified SLC displays were presented in random order once, with the black background either on the left or on the right side of the display. The position of the Munsell scale was instead fixed starting with 2.0 on the left side. There were no time constraints, however participants were instructed to perform matches as quickly as possible. When matches for both targets of one configuration were performed, the experimenter asked participants to close their eyes while a new configuration was positioned. The experiment lasted in average 20 minutes.

Results

Mean matching results for all targets are plotted in Figure 2. In all the analysis that follow, factors *B* and *T* always induced significant effects ($p<.0001$), as expected; we will therefore omit reporting such main effects when we illustrate the results. A 2×3×2×2 ($B \times R \times TC \times T$) ANOVA for repeated measures was conducted on the data; factors Ramp (*R*) and Target Combination (*TC*) produced both significant main effects: *R*, $F_{2,68}=7.31$, $p<.005$; *TC*, $F_{1,34}=7.52$, $p<.01$. Considering two way interactions, only $B \times R$ and $B \times TC$ were significant: respectively, $F_{2,68}=138$, $p<.0001$; $F_{1,34}=7.23$, $p<.05$. A series of one sample t-tests for all mean matches were carried out to verify in which cases mean matches were not statistically distinguishable from the actual reflectance of the target: a star in Figure 2 indicates that a matched reflectance was not statistically distinguishable from its actual reflectance on the Munsell scale.

To better visualize our data, we also grouped them on the basis of factors *TC* and *T*. This determines four groups of data: same targets 6.0, same targets 6.5, and two groups for different targets. Mean matching values for these data groups are graphed as matched reflectance in Figure 3; each graph also shows data for classic SLC displays collected within a different empirical study, however conducted with the same set-up employed here, including the use of targets 6.0 and 6.5, and *TC* as nuance variable (Zavagno et al., 2011).

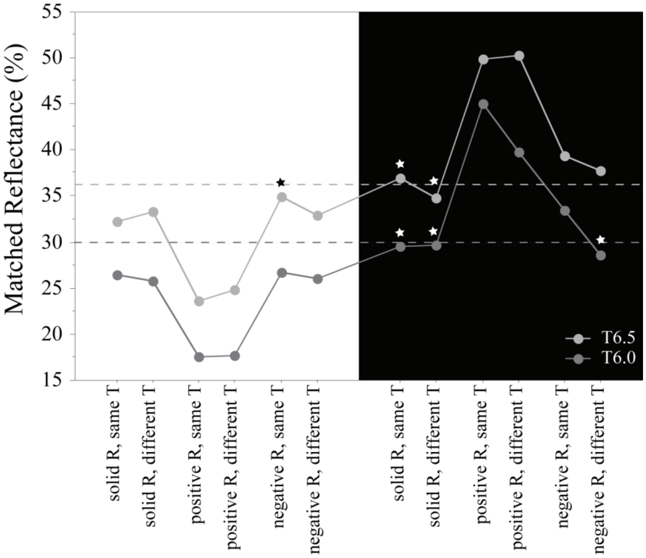


Fig. 2. Mean matching results plotted for each target in terms of matched reflectance. The split white-black background reflects the white-black backgrounds of the configurations. R stands for ramps (positive, negative, or solid, see Figure 1b-d). Same T stands for a configuration in which targets are photometrically equal. Dashed lines indicate the actual reflectance for targets 6.0 and 6.5. Stars above data points indicate mean matches not statistically distinguishable from actual target values.

In the ANOVA we conducted on all data, factor *TC* was found to produce significant effects. This came as a little surprise, since in experiments with 'classic' SLC displays (Figure 1a; data on the left side of the graphs in Figure 3), which purpose was to study the robustness of Achromatic Munsell scales with reference to background changes (Zavagno et al. 2011), the same experimental set-up was used and *TC* did not determine significant main effects. From Figure 2, however, the effects seem to be confined to targets 6.0 seen against a black background in same-different configurations. To test this intuition we ran two separate 2x3x2 (*B*×*R*×*TC*) repeated measure ANOVA's on the data distinguished by factor *T*. The analysis on data for T6.0 revealed a significant effect for *TC* ($F_{1,\ 34}$=10.77, p<.005), but not for *R* (p=.098), while the interactions *B*×*R* and *B*×*TC* are both significant (respectively $F_{2,68}$=129.8, p<.0001, and $F_{1,34}$=8.43, p<.01). The analysis on data for T6.5 show, instead, opposite results: factor *TC* did not induce significant effects (p=.4), while factor *R* did ($F_{2,68}$=7.58, p<.005); only the interaction *B*×*R* induced significant effects: $F_{2,68}$=94.3, p<.0001. Paired t-tests were therefore conducted between targets of identical reflectance, background and ramp, but different for *TC*. With reference to target 6.0, only means for targets on black backgrounds with positive and with negative ramps were statistically different: respectively t_{34}=2.53, p<.05 and t_{34}=3.31, p<.005 (see the dark grey curve on the black side of Figure 2). With reference to target 6.5, none of the paired t-tests revealed significant differences.

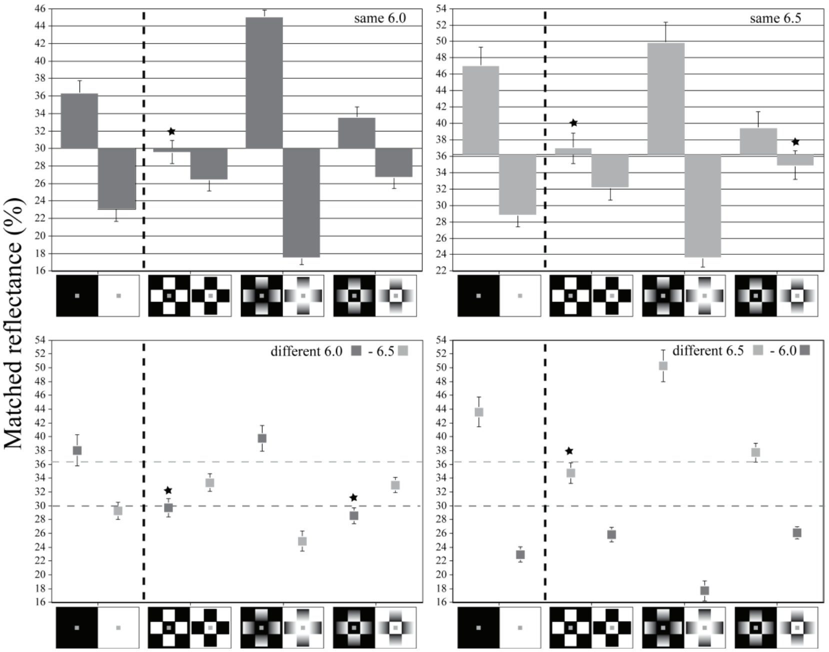


Fig. 3. Results grouped by factor *TC*. Top: results for configurations in which targets shared the same reflectance. Bottom: results for configurations with targets of different reflectance; dashed lines stand for the actual reflectance of targets 6.0 and 6.5. Stars indicate a non-significant difference between matched and actual reflectance (one-sample t-tests). On the left of each graph, separated by a vertical dashed line, are represented data for classic SLC collected within another study (16 observers) but with the same experimental set-up (Zavagno et al., 2011): such data is only used as baseline to understand how SLC increases or decreases in relation to the type of ramp.

On the two groups of data 'same targets' (Figure 3, top graphs) we conducted 3×2 (*R*×*B*) ANOVAs for repeated measures. For T6.0, the ANOVA showed that factor *R* induced significant effects ($F_{2,68}$=3.67, p<.05); also the interaction *R*×*B* was significant (F2, 68=109.6, p<.00001). For T6.5, the ANOVA showed similar results: *R*, $F_{2,68}$=3.13, p<.05; *R*×*B*, $F_{2,68}$=65.55, p<.00001. Finally, paired t-tests were carried out to compare all means respectively for T 6.0 and T6.5 within all same configurations. With regards to T6.0, only in two cases matches for targets were found to be statistically undistinguishable: *white B+solid R* and *white B+negative R* (p=.85), and *black B+solid R* and *white B+negative R* (p=.11). The same pattern of results was found with paired t-tests for T6.5: *white B+solid R* and *white B+negative R* (p=.1), and *black B+solid R* and *white B+negative R* (p=.36).

By comparing our data with the data collected by Zavagno et al. (2011) for classic SLC, we find that with configurations TC=same with T6.0, the magnitude of the effects measured as the mean distance expressed in Munsell units between target matches are approximately the following: 1.25 for classic SLC (Zavagno et al., 2011); 0.25 for solid ramps; 2.5 for positive ramps; 0.5 for negative ramps. With reference to configurations with TC=same with T6.5, the magnitudes of the effects are: 1.25 for classic SLC; 0.25 for solid ramps; 2.25 for positive ramps; 0.25 for negative ramps.

Discussion

By looking at the graphs displayed in Figure 3, a first consideration is that what ever you put inside a SLC display appears to affect the way targets appear, even if the additional elements are not adjacent to the target. In reality, for what we know, the story is a bit more complex. For instance, though global background sizes were equivalent across experiments, our cross configurations seem to limit a target's background to less than 1/6 of the total background. However, in spite of the literature that shows the effect of inducing field size on a target's lightness (the bigger the field, the stronger the induction: Diamond 1955), a target's lightness seems to depend more on the photometric characteristics of surrounding areas than on background size per se. When targets are surrounded by non-contiguous positive luminance ramps, the difference between targets is almost double the difference found with classic SLC. With negative ramps, instead, the effect of simultaneous lightness contrast is about half the effect found with classic SLC. In both cases, increments and decrements in the magnitude of simultaneous contrast are about equivalent for both backgrounds. The story is slightly different when it comes to the 'solid ramps' we used as control stimuli. These configurations share a common photometric feature with negative ramps: the contrast ratio between the edges of the squares facing the target and the target's background are the same. Nevertheless, on black backgrounds the contrast effect of the background on the target is statistically null, for both configurations same and different. This finding is predicted by the Anchoring theory (Economou et al., 1999), according to which the white squares surrounding the targets background become the anchor for lightness scaling, thus dethroning the target from its role of anchor in the local framework. However, it is not clear what the same theory would predict for configurations with luminance ramps (positive or negative). From our data, such ramps do not seem to be lightness features that can affect hypothesized anchoring processes. We believe, instead, that luminance ramps are key information for brightness perception (Zavagno & Daneyko 2008; Zavagno, Daneyko & Sakurai 2011), which in turn affects lightness computations. For instance, with negative ramps on the white background, luminance ramps would inform about a depression in illumination over the target; this in turn appears brighter than it does in classic SLC displays, thus reducing the contrast effect.

References

Agostini T., & Galmonte A. (2002). A new effect of luminance gradient on achromatic simultaneous contrast. *Psychonomic Bulletin and Review, 9,* 264-269.

Diamond A.L. (1955). Foveal simultaneous brightness contrast as a function of inducing field area. *Journal of Experimental Psychology, 50*, 144-152.

Economou E., Zdravkovic S., & Gilchrist A. (2007). Anchoring versus spatial filtering accounts of simultaneous lightness contrast. *Journal of Vision, 7*(12):2, 1–15, http://journalofvision.org/7/12/2/, doi:10.1167/7.12.2.

Kanizsa G. (1979). *Organization in vision. Essays on Gestalt Perception*. New York: Praeger.

Zavagno D. (1999). Some new luminance-gradient effects. *Perception, 28*, 835-838.

Zavagno D., & Daneyko O. (2008). When figure-ground segmentation modulates brightness: The case of phantom illumination. *Acta Psychologica*, *129,* 166-174. DOI: 10.1016/j.actpsy.2008.05.011

Zavagno D., Daneyko O., & Agostini T. (2011). Measuring the meter: On the constancy of lightness scales seen against different backgrounds. *Behavior Research Methods, 43,* 215-223.

Zavagno D., Daneyko O., & Sakurai K. (2011). What can pictorial artefacts teach us about light and lightness?. *Japanese Psychological Research, 53,* 448-462.

Consequential Science
Psychophysical Technologies and Their Implementations

Eugene Galanter & Patricia Hannan

NWEA | Columbia University | Playwisely

An experimental science like psychophysics is often justified, and justifies itself, by the application of its methods to practical affairs. So Fechner's early technological innovations, his "psychophysical methods," that are now adumbrated in elementary textbooks of psychology, were adopted by ophthalmologists, audiologists, and others interested in prescribing emollients for human sensory limitations. They also helped the design and implementation of prosthetic devices based on the psychophysics of these perceptual functions. Advantage—psychosensory methods—they improved the perceptual experience of these folks.

During World War II, psychophysics on the response side, played a major role in training, serving to enhance and extend psychomotor skills of those entrusted with high- (and often low-) speed motor controls on ships and planes, as well as judgmental requirements of high cost decisions. Training techniques in riflery, air-to-air combat, radar judgments, and other complex human performance, demonstrated once again the importance of psychophysical demands on human judgment and competent performance.

These applications of psychophysical methods and knowledge clearly extend themselves to human development, where psychosensory and psychomotor skills are acquired, adjusted, and made part of both quantitative and linguistic mentation. The extension of intrinsic skills by appropriate psychophysical events suggests the application of various forms of remediation, often termed instruction, learning, interacting, and socializing.

We now demonstrate several ways that our discipline can extend itself with practical consequences to the nurturing, education, and socialization of our children. These implementations of our technologies can once again place the science of our field at the forefront of psychological applications to improve the human condition.

There follows a series of audio-video demonstrations of psychophysical applications to demonstrate how psychophysical techniques and procedures can enhance, improve, and ameliorate early limitations of various phases of human development that are bounded by post-natal constraints. You may also see them at: www.PlayWisely.com/research.htm.

The birth and post-natal development of an infant, and the furnishing of an infant's mind requires that the child confronts two profound external demands: First the existence and acknowledgement of other minds within the mind of that child, and then understanding the nature, structure, and the intrinsically limiting character of the physical world.

How did all this start? The early 19th century accepted *confidence in science* to create the right climate. Although often painted as a prim Victorian century, this century is better seen as a time of closing on the truth. Look at the first quarter, not politically, that is well known; Napoleon on St. Helena was dying, Brazil, Argentina, and Peru became independent, Mexico was proclaimed a republic. The Monroe doctrine was to be announced, Liberia was founded, Greece began a struggle for independence, George IV got his cabinet to institute divorce proceedings against Caroline of Brunswick, leading to wilding in the streets of London.

But look beneath the surface at ideas and techniques: In Copenhagen, Ørsted discovered that an electric current twisted a magnetic pole around its conductor. In England, Faraday showed that such a wire would rotate around the pole of a magnet. Wow, the electric motor was born.

But there was more of consequence. Robert Fulton's *North Rive*r *Steam Boat,* popularly known as the "Clermont" went up the Hudson to Albany in 1807. Laënnec invented the stethoscope, Schleiden formulated the cell theory of physiology, Champollion deciphered hieroglyphics, and Lyell established modern geology and the recovery of fossils and fuel. In 25 more years came the central organizing principles of physics and biology; e.g., Helmholtz's conservation of energy, evolutionary statistics by Darwin, modern chemistry that made colored clothing, and Edison who installed the generators and power lines that would brighten New York and London.

So where are *we* in this consequential march from science to life? Psychotherapy and mental health, counseling and stronger marriage, learning theory and teaching, even glasses and hearing aids that fail to move beyond the first two derivatives. Not a day passes that in many newspapers we read of new solutions to educational failure. Try single genders, single students, better assessments for management. Do we need a war to meet the psychological issues that our science can elucidate; hopefully no. Research applications as in physics and biology are only a step away; but our "psychological engineering," are kilometers apart. Without demonstrable instantiations of our psychological understanding, we must stand aside and permit unfounded and often foolish assertions to represent our discipline.

If you believe this view is mistaken, let's have a drink at 6:00, and talk it over.

ACOUSTIC CORRELATE OF PHONOLOGICAL SONORITY IN BRITISH ENGLISH

Yoshitaka Nakajima*, Kazuo Ueda*, Shota Fujimaru*, Hirotoshi Motomura*, Yuki Ohsaka*
*Kyushu University, Fukuoka, Japan
nakajima@design.kyushu-u.ac.jp

Abstract

Sonority or aperture proposed in linguistic literature can be considered a kind of subjective measure specific to speech perception. Vowels have high sonorities corresponding to a linguistic fact that they can be nuclei of syllables, and fricatives and stops have low sonorities. In order to understand how sonority is perceived, we attempted to find an acoustic dimension on which we could construct a psychophysical scale of sonority. We applied the multivariate analysis method as in Ueda et al. (2010, Fechner Day, Padua) to spoken sentences in British English collected in a commercial database, in which phonemes were segregated and labeled. The speech signals went through a bank of critical-band filters, and the output power fluctuations were subjected to factor analysis. The three factors as in our previous study appeared. The analyzed phonemes were classified into three categories, i.e., vowels, sonorant consonants, and obstruents. These categories were represented well in the Cartesian space whose coordinates were the factor scores of the above factors. One of the factors located around 1000 Hz was highly correlated with sonority or aperture.

The concept of the syllable is important in accounting for why the temporal order of phonemes is often guided by a set of rules or constraints (e.g., Spencer, 1996; Prince & Smolensky, 2004). This study was originally an attempt to find a psychoacoustic basis to understand how syllables were formed. We were interested in whether the three-factor representation of speech sounds as in Ueda et al. (2010) could be related to their phonological categories. Vowels are often represented in a formant map, and this helps a lot to understand how vowels are articulated and perceived. We took a further step: We took up spoken sentences in British English, and tried to draw a map of vowels, including diphthongs, and consonants put together on a purely acoustic basis. Most consonants and all diphthongs are characterized by systematic spectral changes in time, and these changes are often important for speech perception. However, we did not take up such spectral changes in the present stage because they seemed too complicated to be reflected in a simple map. We thus took up the following strategy. We specified representative factor scores for each observed phoneme as a first step. We then observed the results to judge whether they reflected essential aspects of syllable formation. If so, then we could take a next step, and to analyze spectral changes within phonemes would be one possible alternative. If not so, then the present analysis would not be promising, and another way should be looked for. Now, the first step resulted in a configuration of phonemes which seemed worth reporting as it was. Thus, we report how the configuration was obtained and interpreted.

We looked for a database of speech sounds in which most phonemes of a certain language appeared, and were segregated and labeled. Fortunately, a commercial database of British English [The ATR British English speech database (Campbell, 1993)] was available to study this issue. This database was developed for speech science research. British English would give us a secure starting point, because its phonology has been described thoroughly in the literature (e.g., Harris, 1994; Spencer, 1996). About 25,000 samples of phonemes uttered by three speakers were available; the sheer amount of data was a great advantage of utilizing this database.

Analysis

We analyzed spoken sentences in British English to extract the three factors to explain part of the power fluctuations in twenty neighboring critical bands. We determined factor scores of each labeled phoneme at the temporal middle of its labeled duration.

Method

Speech samples The database comprises two-hundred English sentences read aloud by three native speakers of British English: two females and one male (There was another male speaker, but there was a technical problem in the labeling data of this speaker, which could not be fixed by the company that released the database). The speech signals were recorded with 12-kHz sampling and 12-bit quantization. Labeling was performed manually; all the phoneme labels were linked to specific periods of speech signals that sounded approximately as indicated if played separately.

The labeling data sometimes did not completely agree with the phonemes indicated in dictionaries expressing subtle differences of the sounds, but such differences were not within our present interest, which was to connect the acoustic and the phonological features of the speech signals. The labeling data thus had to be modified in some cases. We preserved the original labeling data of the database as far as possible, but incoherent cases were omitted. There were

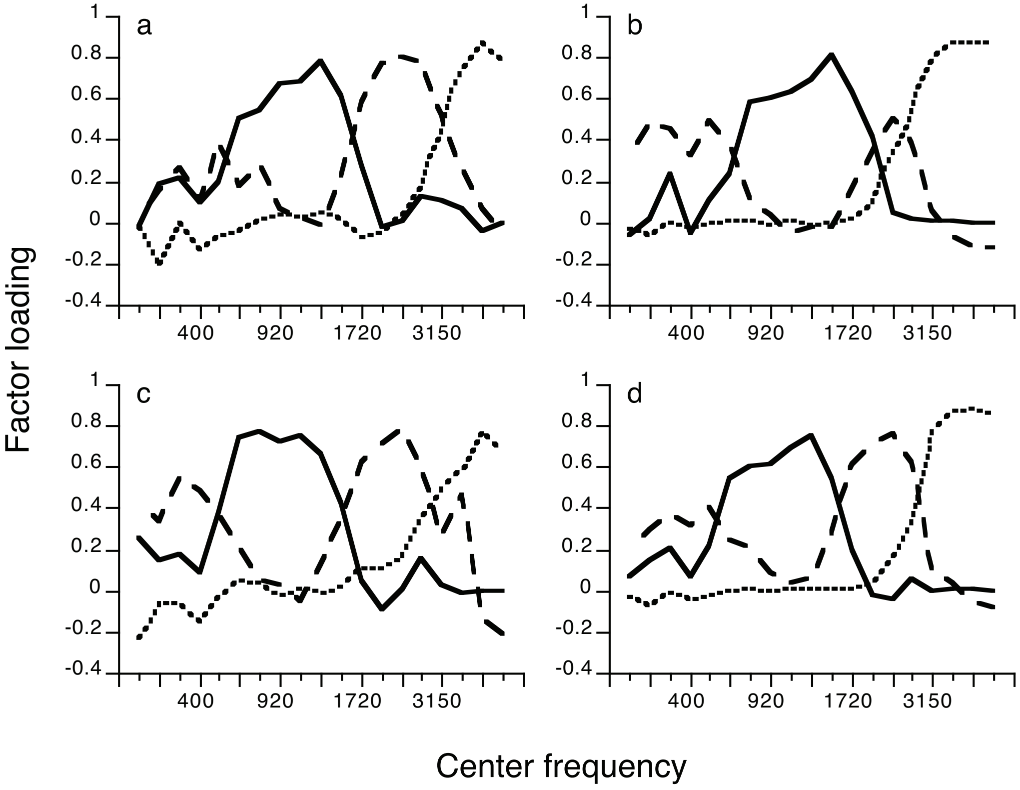


Figure 1 Obtained factors. Panels (a) and (b) show the results for two female speakers, respectively, and panel (c) shows the results for a male speaker. The results for all the speakers are combined in panel (d).

cases, although few, in which voiced and unvoiced phonemes were exchanged in the labeling data; these cases were not omitted, and the phoneme labels in the database were utilized. Some phonemes were separated into more than one period in the database. One representative period was chosen in such a case; a closure part was always omitted from further analyses, even when it was the only part representing a phoneme, because closure parts had almost no sound energy to differentiate them from each other. When a vowel was separated into a main part and a transient nasalized part, the main part was chosen. If a stop consonant was still separated into more than one period, the period in which a stronger aspiration was indicated in the label was chosen. Out of the 31 663 labeled periods in the database, 6 754 periods were omitted in the above screening.

Procedure All the speech signals were analyzed as in Ueda et al. (2010). Power fluctuations were derived from 19 critical-band filters (Zwicker & Terhardt, 1980). The lowest critical band was 50-150 Hz; spectral components below 50 Hz were neglected because no substantial parts of speech signals were to be expected in this range. The 19th, i.e., the highest, critical band was 4800-5800 Hz. The derived power fluctuations were submitted to principal component analyses, and factors were determined by varimax rotation.

Each labeled period of the database was then connected to a set of factor scores. The outputs of the critical-band filters were utilized in this calculation; the output powers at the temporal middle of the period were converted into factor scores by applying the factor loadings obtained in the above analysis.

Thus, all the labeled phonemes as screened above were represented in a Cartesian space of the factor scores, i.e., a factor space. Finally, each English phoneme was represented by a single point in the space; the sets of factor scores were averaged for each English phoneme. A configuration of the English phonemes was thus obtained by a purely acoustic analysis; exactly speaking, the labeling data in the speech database were the only linguistic information utilized in the present analysis.

Results

The varimax rotation led to three factors to be related to four frequency ranges as in Figure 1. The speech signals of the three speakers were combined and submitted to a single analysis

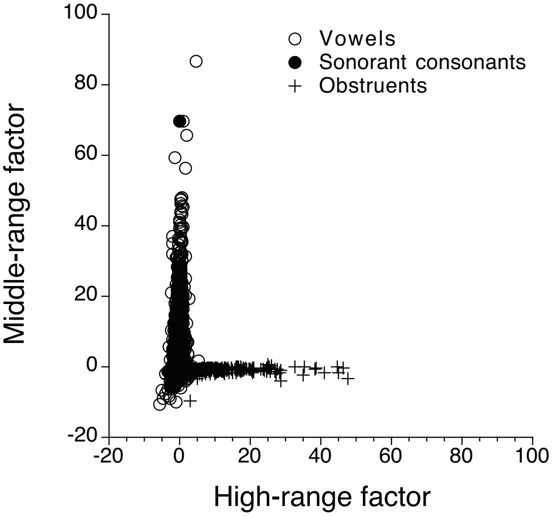


Figure 2 An L-shaped distribution of labeled phonemes on a plane of the high-range and the middle-range factor.

(Figure 1d), and the speech signals of individual speakers were also analyzed separately (Figures 1a-c). The cumulative contributions of the three factors were 41-45% in all these analyses. The following three factors appeared: the *high-range factor* appeared above 3200 Hz, the *middle-range factor* around 1100 Hz, and the *exterior factor* in two separate frequency ranges around 300 and 2200 Hz. The results were consistent with our previous results (Ueda et al., 2010).

The labeled phonemes were represented in a three-dimensional factor space corresponding to the above factors. Their configuration was characteristically L-shaped on the plane of the high-range and the middle-range factor (Figure 2). The high-range factor seemed to be related only to obstruents (fricatives, affricates, and stops), and the middle-range factor only to vowels and sonorant consonants (glides, liquids, and nasals). Although vowels and sonorant consonants shared a substantial area, their distributions differed; vowels could take higher scores of the middle range factor, and the distribution of sonorant consonants were often limited to an area that was shared by vowels but related to smaller values of the middle-range factor scores.

The high-range factor and the middle-range factor thus could be expressed as a single dimension by observing the origin of the space from a viewpoint in the space in which both factors took positive values (Figure 3). Vowels and sonorant consonants were separated clearly from obstruents.

Each English phoneme was represented by the average position (center of gravity) of all the labeled samples in this space (Figure 4). In this presentation, the phonemes were clearly separated into three categories: vowels, sonorant consonants, and obstruents.

Discussion

The obtained configuration of the phonemes seems to have a close relationship with English syllable formation. Most English vowels can be nuclei of stressed syllables, and they are distributed in the upper right part of the configuration. On the contrary, obstruents can never be syllable nuclei, and they are distributed in the lower left part. Sonorant consonants, which are considered more vowel-like than obstruents, are located near the middle of the configuration; note that /l/ and /n/ can be nuclei of unstressed syllables.

When the factor scores of the high-range and the middle-range factor are near or below

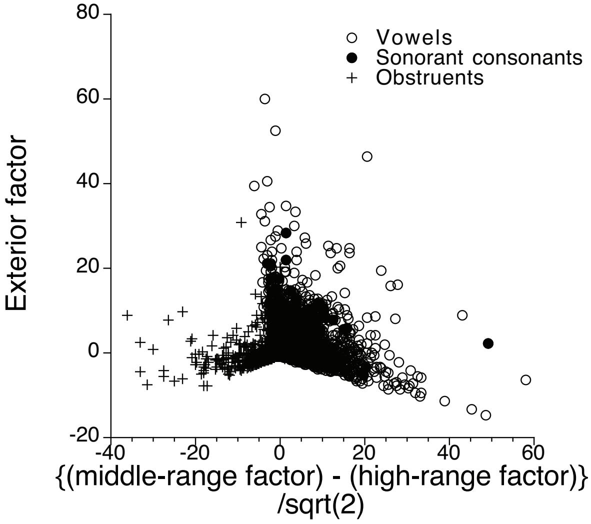


Figure 3 Distribution of labeled phonemes in the factor space.

zero, as is observed near and above the origin in Figure 4a, the factor score of the exterior factor can change in a wider range; vowels are located in the upper part, and obstruents in the lower part.

The nature of phonemes related to the gradual change through vowels, sonorant consonants, and obstruents is called *sonority*, or *aperture*, in linguistics. Typically, Spencer (1996) proposed the following sonority scale, in which larger numbers indicate higher sonority: 6) vowels, 5) glides, 4) liquids, 3) nasals, 2) fricatives/affricates, 1) plosives (stops). The English

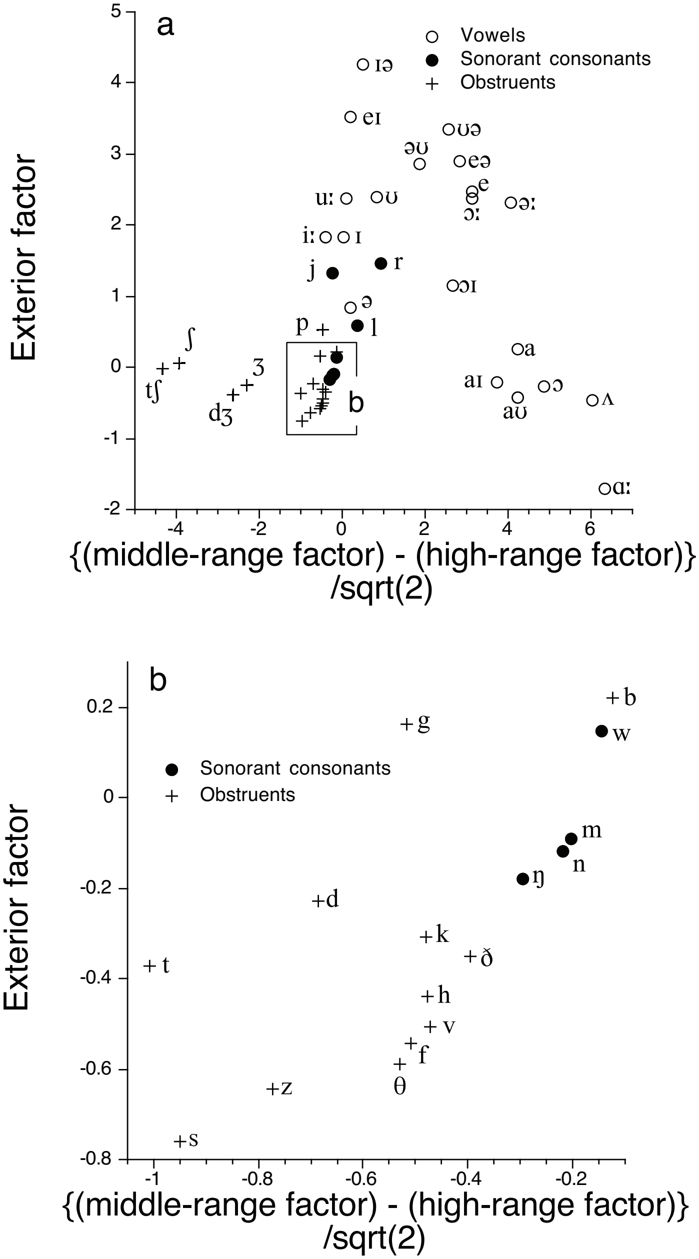


Figure 4 A map of English phonemes based on an acoustic analysis.

phonemes represented in the above factor space can be related to this sonority scale. The Spearman's rank correlation coefficient between the middle-range factor score and the sonority scale value was as high as 0.83 (N=44; $p<0.05$). The other two factors also showed significant rank correlations: -0.45 for the high-range factor ($p<0.05$) and 0.54 for the exterior factor ($p<0.05$). For a classic example, de Saussure (1959) proposed a similar scale of aperture (in an appendix, which is rarely quoted), and the aperture value showed a significant rank correlation (0.83; N=31) with the middle factor score. Considering the factor loadings related to the center frequencies of the critical-band filters (Figure 1), sonority is very likely to be a subjective dimension corresponding, psychophysically, to the acoustic power below 2500 Hz; the acoustic power around 1000 Hz seems especially important, creating the impression of stressed syllables. The acoustic power above 4000 Hz may work to suppress sonority, making the impression of syllable boundaries clearer.

General Discussion

Sonority, which has been a purely linguistic concept created mainly to understand phonological rules or constraints especially in syllable formation, is now a concept that seems useful in computational and psychophysical approaches to auditory perception. The middle-range factor located around 1000 Hz can be a first approximation of a physical property to be put against sonority in psychophysical studies. It is also important that frequency components above 4000 Hz may be playing important roles in speech perception. Such components are neglected in the present telephone communication, and we are now developing a system in which the information carried by the high frequency components can be transmitted via a lower frequency channel.

Acknowledgments. Ger Remijn, Valter Ciocca, Mark Elliott, and Peter Howell gave us valuable suggestions. This research was supported by grants to YN and KU from the Japan Society for the Promotion of Science (19103003, 20653054) and from Kyushu University (an A type project; The Kyushu University Project for Interdisciplinary Research of Perception and Cognition).

References

Campbell, N. (1993). The ATR British English speech database. *Technical Report*, ATR Interpreting Telephony Research Labs.

de Saussure, F. (1959). *Course in General Linguistics*. (Baskin, W., Trans.). McGraw-Hill, New York (Original work published 1916).

Harris, J. (1994). *English Sound Structure*. Blackwell, Oxford.

Prince, A. & Smolensky, P. (2004). *Optimality Theory: Constraint Interaction in Generative Grammar.* Blackwell, Malden.

Spencer, A. (1996). *Phonology: Theory and Description*. Blackwell, Oxford.

Ueda, K., Nakajima, Y., & Satsukawa, Y. (2010). Effects of frequency-band elimination on syllable identification of Japanese noise-vocoded speech: Analysis of confusion matrices. *Fechner Day 2010*, 39-44.

Zwicker, E. & Terhardt, E. (1980). Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. *Journal of the Acoustical Society of Amarica, 68*, 1523-1525.

PERCEPTUAL ROLES OF DIFFERENT FREQUENCY BANDS IN JAPANESE SYLLABLE IDENTIFICATION

Kazuo Ueda and Yoshitaka Nakajima
Department of Human Science and Center for Applied Perceptual Research
Kyushu University, 4-9-1 Shiobaru Minami-ku, Fukuoka 815-8540, Japan
Kiyoto Noguchi and Yuichi Satsukawa
Department of Acoustic Design, Kyushu University, 4-9-1 Shiobaru Minami-ku
Fukuoka 815-8540, Japan
ueda@design.kyushu-u.ac.jp *nakajima@design.kyushu-u.ac.jp*

Abstract

Ueda et al. [(2010). Fechner Day 2010, Padua.] indicated that speech information could be essentially transmitted by the power fluctuations in four frequency bands. We aimed at clarifying the roles of these frequency bands in Japanese speech perception in V/CV syllable identification. We first performed factor analyses of power fluctuations of critical-band-filtered speech, and obtained four frequency bands as in the previous research. The speech was a set of V/CV patterns uttered by a male and a female speaker. The speech patterns were converted into noise-vocoded speech so that only the power fluctuation in each frequency band was preserved. There were also patterns in which one of the frequency bands was eliminated resulting in a spectral gap. Eliminating the lowest band (50-570 Hz) crucially deteriorated perceptual differentiation between voiced and unvoiced consonants. Eliminating the second lowest band (570-1850 Hz) interfered vowel identification turning almost all vowels into /i/. The roles of the other frequency bands were not obvious, but their temporal relationships with the lowest band was suggested to play a role.

The aim of the present investigation was to clarify how elimination of one of four frequency-bands in noise-vocoded speech of Japanese V/CV syllables affected identification, in terms of the amount of information transmitted obtained from confusion matrices.

Ueda, Nakajima, and Satsukawa (2010) extracted three factors of power fluctuations common to eight languages. These factors had four non-overlapping mounds in factor loadings, and the frequency range of speech were divided into four frequency-bands separated by boundaries around 500, 1700, and 3300 Hz; each band centered around one of the mounds. Nakajima, Ueda, Fujimaru, Motomura, and Ohsaka (2012) examined the correspondence between factor scores and phonemic labels in British English, and found that these three factors clearly differentiated three phonemic categories: vowels, sonorant consonants, and obstruents. These factors were correlated to *sonority* or *aperture* (e.g., de Saussure, 1959; Selkirk, 1984; Spencer, 1996).

Another way to explore potential correspondence between those factors and phonemic perception would be to simplify the three factors into power fluctuations in the four frequency bands by utilizing noise-vocoded speech (e.g., Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Smith, Delgutte, & Oxenham, 2002; Sheldon, Pichora-Fuller, & Schneider, 2008; Roberts, Summers, & Bailey, 2011), and observing the effects of manipulating these frequency bands on perception. Miller and Nicely (1955) and Benkí (2003) estimated the amount of information transmitted through human listeners, based on confusion matrices of consonant identification. Employing a similar analysis method,

Table 1 Experimental conditions.

Condition	Number of bands	Eliminated band (Hz)
1: Original	-	-
2: 20-band noise-vocoded	20	-
3: 4-band noise-vocoded	4	-
4: 1st-band eliminated	3	50-570
5: 2nd-band eliminated	3	570-1850
6: 3rd-band eliminated	3	1850-4000
7: 4th-band eliminated	3	4000-7000

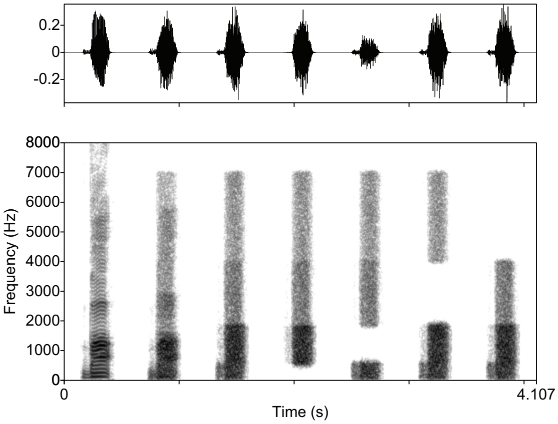


Figure 1 Examples of stimuli based on a Japanese CV-syllable, /ba/. The upper panel shows the waveforms and the lower panel the corresponding spectrograms. In each panel, from left to right, the experimental conditions 1 to 7 in Table 1 are exemplified.

we eliminated one of the frequency-bands in noise-vocoded speech of Japanese V/CV syllables, and estimated the amount of information transmitted by each frequency band.

Experiment

Method

Experimental conditions and stimuli Experimental conditions are listed in Table 1. A total of 101 Japanese V/CV syllables uttered by a male and a female speaker was edited and extracted from "ATR Digital Speech Database Set A" (20-kHz sampling and 16-bit quantization). In condition 1, the edited recordings of Japanese V/CV-syllables were played. Noise-vocoded speech stimuli were prepared for conditions 2-7. Critical bands (e.g., Zwicker & Terhardt, 1980) were adopted over the range of 50-7000 Hz in condition 2. Examples of stimuli are shown in Figure 1.

Table 2 Japanese consonant classification utilized in the analysis.

Consonant classification	Phonemes
Voicing	
Voiced	/b, d, g, z, m, n, N, ɾ, y, w/
Unvoiced	/p, t, k, s, h/
Manner of articulation	
Stops	/p, t, k, b, d, g/
Fricatives	/s, z, h/
Nasals	/m, n, N/
Flap	/ɾ/
Semivowels	/y, w/

Participants All the participants were normal-hearing Japanese speakers. Eight listeners (5 males and 3 females, mean age = 26, ranged from 21 to 45) participated in experiment sessions where stimuli based on syllables uttered by a male speaker were presented. Ten other listeners (5 males and 5 females, mean age = 23, ranged from 22 to 25) participated in experiment sessions where stimuli based on syllables uttered by a female speaker were presented.

Procedure The stimuli were diotically presented to each listener through headphones (STAX SR-303). Each stimulus allotted to three trials in random order for each listener. Within each trial, presentation of the same stimulus was repeated three times in succession with an inter-stimulus-interval of about 2 s. The sound pressure level was adjusted so that the peak level of a vowel /a/ became 78 dB A (Fast). The listeners were instructed to identify what they heard by selecting an appropriate button on a screen as far as possible. When they could not identify a stimulus, they were instructed to press a “?” button, or to write down what they heard on an answer sheet.

Results

The listeners’ responses were collected as confusion matrices of 101 × 101. The responses that fell outside the matrices were neglected, because they were rare (0.08% in the highest condition). Relative amount of information transmitted was calculated according to the method utilized by Benkí (2003). Separate information channels were assumed for vowels and consonants in the calculation. Vowels were classified into five categories, /a, e, i, o, u/. The classification utilized for consonants was summarized in Table 2. Figures 2 and 3 show the relative amounts of information transmitted. Figure 2 shows three effects on the relative amounts of information transmitted: the effects of noise-vocoding, number of bands, and elimination of specific frequency bands. Noise-vocoding generally reduced the amounts of information transmitted, except in the 20-band condition in vowel distinction (Fig. 2a). Reducing the number of bands from 20 to 4 resulted in reducing the amounts of information transmitted, except in voicing distinction (Fig 2b). Eliminating a specific frequency band, especially the second lowest band in vowel distinction and the lowest band in voicing distinction, largely reduced the amounts of information transmitted (Fig. 2a,b). Figure 3 shows similar effects concerning noise-vocoding and the number

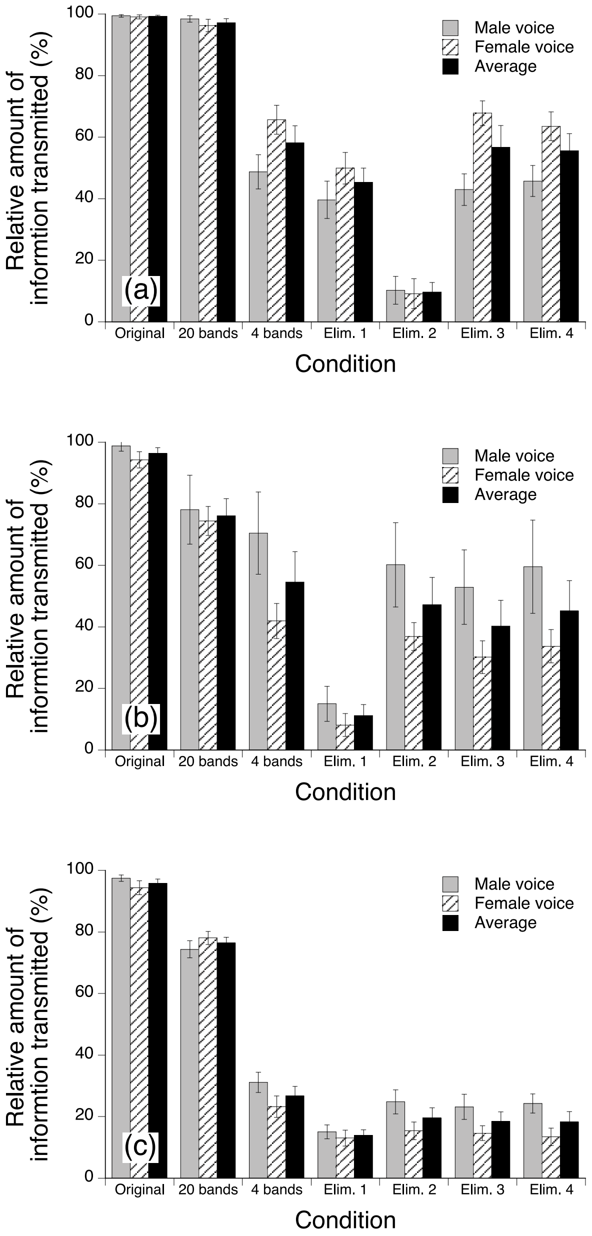


Figure 2 Relative amounts of information transmitted, calculated in terms of identifying (a) vowels, (b) voicing, and (c) manners of articulation. The error bars represent 95% confidence intervals.

of bands, whereas the effect of eliminating specific frequency bands differed very much between consonant classification categories. The amounts of information transmitted for fricative distinction reduced sharply when the lowest frequency band was eliminated (Fig. 3b). A nonparametric multiple comparison test (Steel-Dwass test) supported these

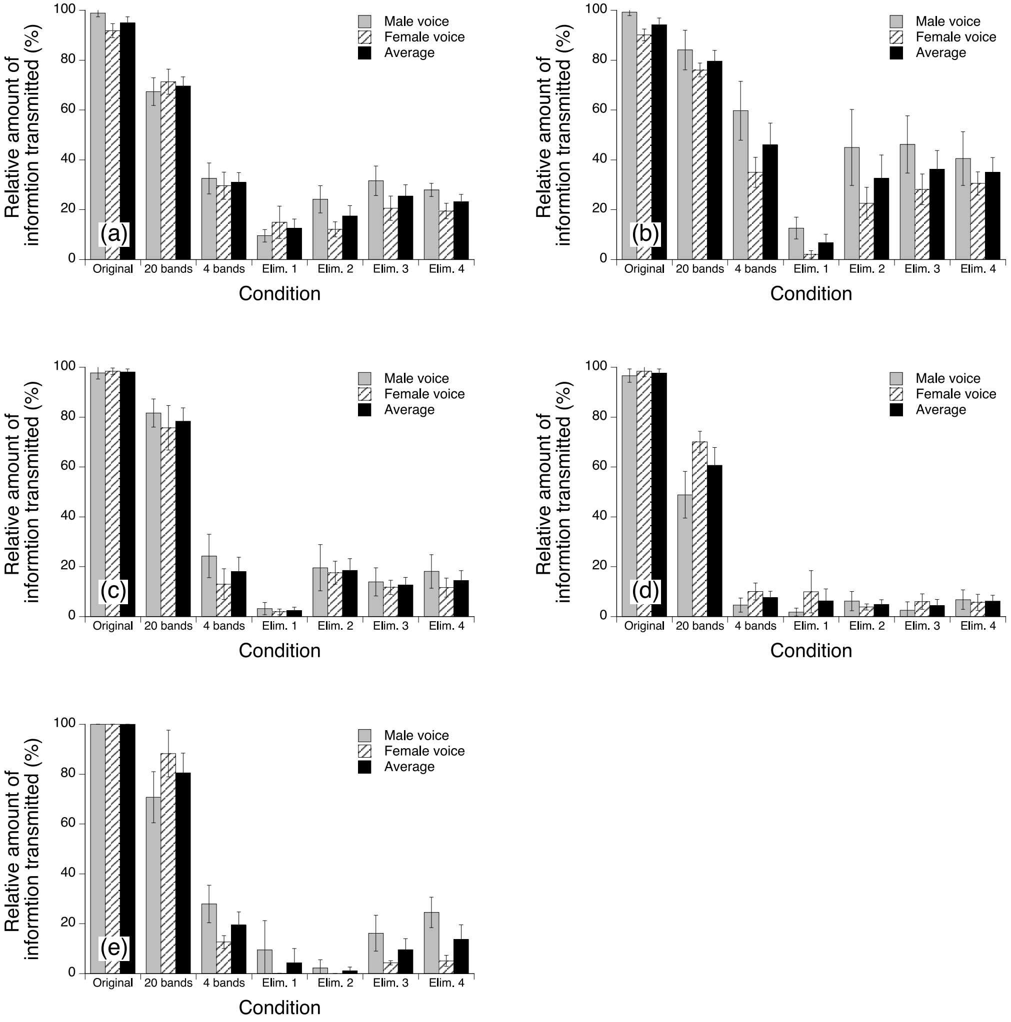


Figure 3 Relative amounts of information transmitted, calculated in terms of identifying categories within manners of articulation: (a) stops, (b) fricatives, (c) nasals, (d) flap, and (e) semivowels. The error bars represent 95% confidence intervals.

observations.

Discussion

The lowest frequency band should not be removed when information concerning voicing of consonants should be conveyed. It is possible that a short forerunning energy in the lowest frequency band preceding a burst, which can be seen in Figure 1, signals a voiced consonant, as a substitute of a voice bar, whereas the absence of the lowest frequency band tends to signal a voiceless consonant. Thus, voicing can be cued without periodicity of voicing. Removing the second lowest frequency band may distort formant structure of vowels, resulting in a fixed formant pattern which resembles that of the vowel /i/.

The roles of the third and the fourth frequency bands were not obvious in the present investigation, but, it is possible that these frequency bands have some roles in phoneme perception in combination with other frequency bands; the amounts of information transmitted to signal fricatives may be related to the temporal relationship between the lowest frequency band and the third or fourth band.

Acknowledgments

This research was supported by Grants-in-Aid for Scientific Research Nos. 14101001, 19103003, and 20330152 from the Japan Society for the Promotion of Science, Kyushu University Interdisciplinary Programs in Education and Projects in Research Development, and a Grant-in-Aid for the 21st Century COE program from the Ministry of Education, Culture, Sports, Science and Technology.

References

Benkí, J. R. (2003). Analysis of English nonsense syllable recognition in noise. *Phonetica*, *60*, 129-157.

de Saussure, F. (1959). *Course in general linguistics (Baskin, W., Trans.).* New York: McGraw-Hill Paperbacks. (Original work published 1916)

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *J. Acoust. Soc. Am.*, *27*, 338-352.

Nakajima, Y., Ueda, K., Fujimaru, S., Motomura, H., & Ohsaka, Y. (2012). Acoustic correlate of phonological sonority in British English. In *Fechner Day 2012, Proceedings of the 28th Annual Meeting of the International Society for Psychophysics.* Ottawa, Canada.

Roberts, B., Summers, R. J., & Bailey, P. J. (2011). The intelligibility of noise-vocoded speech: spectral information available from across-channel comparison of amplitude envelopes. *Proc. Royal Soc. B*, *278*, 1595-1600.

Selkirk, E. (1984). On the major class features and syllable theory. In M. Aronoff & R. T. Oehrle (Eds.), *Language Sound Structure: Studies in Phonology Presented to Morris Halle by His Teacher and Students* (p. 107-136). Cambridge, MA: MIT Press.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303-304.

Sheldon, S., Pichora-Fuller, M. K., & Schneider, B. A. (2008). Effect of age, presentation method, and learning on identification of noise-vocoded words. *J. Acoust. Soc. Am.*, *123*, 476-488.

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, *416*(7 March 2002), 87-90.

Spencer, A. (1996). *Phonology: Theory and Description*. Oxford: Blackwell.

Ueda, K., Nakajima, Y., & Satsukawa, Y. (2010). Effects of frequency-band elimination on syllable identification of Japanese noise-vocoded speech: Analysis of confusion matrices. In *Fechner Day 2010, Proceedings of the 26th Annual Meeting of the International Society for Psychophysics* (p. 39-44). Padova, Italy.

Zwicker, E., & Terhardt, E. (1980). Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. *J. Acoust. Soc. Am.*, *68*, 1523-1525.

French and English rhythms are perceptually discriminable with only intensity changes in low frequency regions of speech

Tsuyoshi Kuroda[†], Simon Grondin[†], Yoshitaka Nakajima[‡], Kazuo Ueda[‡]
[†]*Université Laval, Québec, Canada,* [‡]*Kyushu University, Fukuoka, Japan*
tsuyoshi.kuroda.1@ulaval.ca

Abstract

The purpose of this study was to determine which frequency band would contribute to discrimination between speech rhythms of French and English. Each trial consisted of two noises with different intensity changes. Each intensity change simulated the one that was derived from a frequency band of recorded sentences of French or English; the band had a center frequency of 350, 1000, 2150, or 4800 Hz. Participants evaluated the rhythm dissimilarity of two noises with an 8-point scale. Two noises were evaluated as more dissimilar when two sentences whose intensity changes were simulated by the noises were in different languages than when they were in the same language. Moreover, this tendency was reduced in 4800 Hz compared with the other bands. This indicates that French and English rhythms are discriminable with intensity changes of low frequency bands, even without any signs of pitch and phoneme.

Languages have their own rhythm. They are categorized according to what linguistic units comprise their rhythm. For instance, French is regarded as a *syllable-timed* language where syllables are regularly uttered while English is regarded as a *stress-timed* language where stressed syllables are regularly uttered (Fant, Kruckenberg, & Nord, 1991). However, the validity of this categorization is not well established by acoustic and psychological studies (see Patel, 2008). This seems due to the fact that there are few studies determining what acoustic property is linked to speech rhythm.

However, Nazzi, Bertoncini and Mehler (1998) demonstrated that newborn infants could discriminate different languages with only prosodic cues. In their experiment, speech stimuli were low-pass filtered to remove semantic cues but keep prosodic cues. Because newborn infants were naïve to semantic aspects of speech, their results indicated that languages could be discriminated even without semantic cues. Moreover, infants could discriminate between English (stress-timed) and Japanese (mora-timed) but not between English and Dutch (stress-timed). This supports the contemporary categorization of language rhythms and indicates that language rhythms are linked to some acoustic properties in low frequency regions of speech.

The method of Nazzi et al. (1998) could work only when infants were employed as participants, while the present study developed a method that could be used for adults to determine what acoustic property would be linked to speech rhythms of French and English. Note that Nazzi et al.'s stimuli included pitch cues because these included fundamental frequencies. However, the present experiment examined whether French and English rhythms could be discriminated with *only* temporal changes of intensity included in speech, and also examined which frequency band would be crucial for discriminating between French and English rhythms. In each trial, two noises were successively presented. These noises had different intensity changes. Each intensity change simulated the one that was derived from a frequency band of recorded sentences of French or English. Participants were instructed to evaluate the rhythm dissimilarity of two noises. Note that the noises did not include any signs of pitch and phoneme of speech. If French and English rhythms are discriminated with only intensity changes included in speech, two noises should be perceived as dissimilar rhythms when one noise is derived from a French sentence and the other from an English sentence.

Method

Participants

Fifteen participants, five males and ten females aged 20-40 years, were recruited. They were students or employees at Université Laval. They consented to their participation by signing a form approved by the institutional ethical committee and received $100 CAN for their participation. Eight participants reported that they were French speakers and seven reported that they were English speakers.

Apparatus and stimuli

Recorded speech sentences which were sampled at 16000 Hz and quantized to 16 bits in an electric database (NTT-AT multilingual speech database 2002, 2002) were used. Four French and four British English sentences were randomly selected from the database for each participant; however, only sentences whose duration was approximately between 1.6 and 2.8 s were selected. Since each sentence was spoken by five males and five females in each language in the database, two males were randomly selected for two of the four selected sentences and two females were randomly selected for the other sentences, i.e., sentence 1 was spoken by male 1, sentence 2 by male 2, sentence 3 by female 1, and sentence 4 by female 2.

The procedure of making stimuli is illustrated in Figure 1. Each sentence passed through a band-pass filter, whose parameter was decided with Bark scale (Scharf & Buus, 1986). The filter always had a range of 2 Bark while its center was located on 4, 9, 14, or 19 Bark. The filter condition is called with a frequency corresponding to a center bark of the filter, i.e., 350, 1000, 2150, and 4800 Hz.

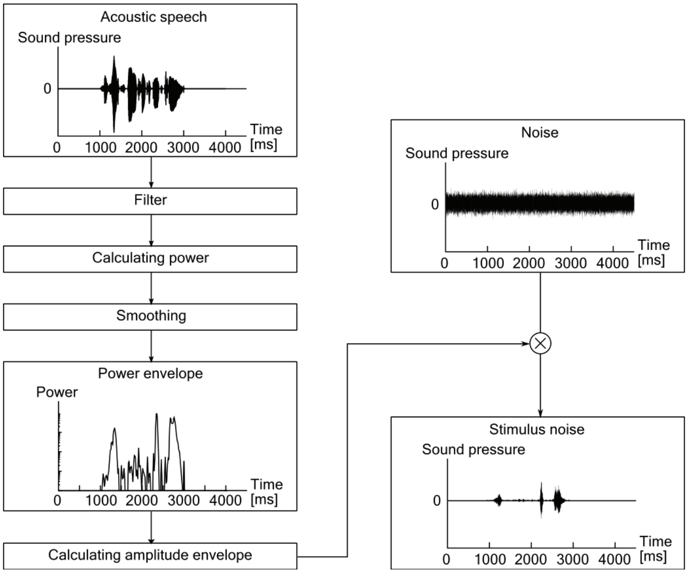


Figure 1. Procedure of making stimuli.

Amplitude (sound pressure) at each sample of the filtered speech was squared to obtain a power envelope. The resulting envelope was smoothed by calculating a moving average with a Gaussian window whose standard deviation was 5 ms. Then, a white noise of 150-7000 Hz was generated and its amplitude was multiplied by a square root of power at each sample of the smoothed envelope; consequently, the noise had the same power envelope as the filtered speech. There were 32 noises (2 languages × 4 sentences × 4 filters) in total.

Stimulus intensity was calibrated in two conditions. In one, the root-mean-square (RMS) level of each noise was calibrated at 25 dB sensation level (SL), i.e., all noises had equal RMS levels (mean-equal condition). In the other, the peak intensity of each noise was calibrated at 35 dB SL, i.e., all noises had equal peak levels (peak-equal condition). The reason why these conditions had different reference levels (25 and 35 dB SL) is because the mean-equal condition makes stimuli even louder than the peak-equal condition if these conditions have equal reference levels. Different observers were allocated to each of the two conditions, i.e., these conditions were between-participants conditions. Four French-speaking and three English-speaking participants were allocated to the mean-equal condition while four French-speaking and four English-speaking participants were allocated to the peak-equal condition.

Procedure

Two noises were successively presented in each trial. Participants were instructed to evaluate the rhythm dissimilarity of two noises with an 8-point scale, where “1” indicates “exactly the same” and “8” indicates “extremely dissimilar.” Scales of around 8 points were used in multidimensional-scaling studies (e.g., Abelson, 1954-55). Participants responded by clicking on a pane on a computer display. They listened to stimuli by clicking on the “play” pane. They were allowed to listen to stimuli only once in each trial, but when listening was disturbed for some specific reason (e.g., yawning or coughing), they could listen to the stimuli again by clicking on the “replay” pane. Two noises were separated by an inter-stimulus interval varied from 2.5 to 3 s randomly.

Because the same noise was not doubly presented within one trial, there were 992 trials ($_{32}P_2$). The experiment took two sessions each consisting of the 992 trials, i.e., participants responded twice for each of the 992 trials. The order of the trials was randomized in each session. Each session was divided into 16 blocks each consisting of 62 trials. Before the beginning of the experiment, threshold intensity for detecting a noise was measured and a practice block consisting of 62 randomly selected trials was carried out. Participants completed the threshold measurement and the practice block in one day while they completed the experimental sessions over 16 days (two blocks per day). Thus, the experiment was completed over 17 days. Each block took about 20 minutes, and a break of a few minutes was taken between blocks in each day. There were two warm-up trials at the beginning of each block and these trials consisted of the stimuli that were to be presented in the last two trials of the block.

Results

Each participant made four responses for each of the 496 noise pairs; the order of two successive noises was collapsed (992 trials ÷ 2 orders). These four responses were averaged. Kruskal’s nonparametric multidimensional scaling was conducted on a dissimilarity (triangular) matrix for each participant (with software of R version 2.14.1). The reason why this analysis was conducted on individual data instead of pooled data is because four sentences were randomly selected for each participant, i.e., participants had different sets of sentences.

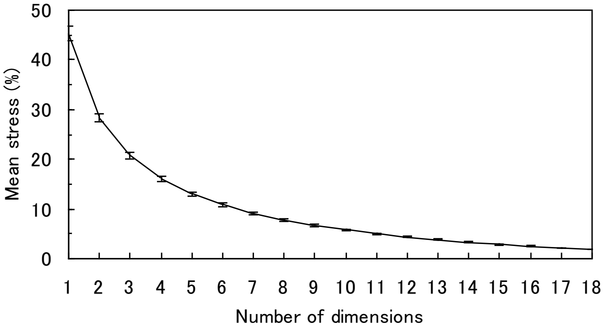


Figure 2. Mean stress in each number of dimensions that were constructed by a multidimensional scaling. Bars are standard error of mean.

The scaling approach has been taken in literatures to visualize relations between stimuli in a space of a few dimensions. Unfortunately, in the present experiment, the *stress* was too high, i.e., the goodness of fit was too poor, to visualize the data with a few dimensions (Figure 2). As an alternative method for approaching the data, the scaling was constructed with as many dimensions as possible; it was constructed with eighteen dimensions where the *stress* was sufficiently low (1.9%). In this approach, the multidimensional scaling was utilized simply for converting the nonparametric dependent variables, i.e., dissimilarity ratings, into the parametric ones; the number of dimensions had to be increased to minimize the stress.

If French and English rhythms are discriminated with power fluctuations in a limited frequency region, a Euclidian distance between French and English sentences, which were located in the 18-dimentional space, should be increased in a specific band relative to the other band conditions. Because there were four English and four French sentences, we calculated 16 Euclidian distances for 16 pairs of English vs. French sentences (4 × 4) in each band condition. These 16 distances were squared and summed up, the resulting value indicating how far apart English and French conditions were in each band in the 18-dimensional space (the *two-language separation* — TS). In addition, 28 Euclidian distances for all 28 pairs of sentences in each band (${}_8P_2$ because there were 8 sentences, i.e, 4 French + 4 English sentences) were squared and summed up, the resulting value indicating how large a divergence between individual sentences was (the *individual-sentence divergence* — ID). Dividing TS by ID gave an index of *relative separation* (RS). Note that RS becomes around .57 if all pairs of sentences lead to equal distances; for example, if all pairs lead to a distance of 1, TS becomes 16 and ID becomes 28, resulting in RS of around .57 (16 ÷ 28). In other words, RS above .57 indicates that two power envelopes (noises) were evaluated as more dissimilar rhythms when these envelopes were obtained from sentences in different languages than when obtained from sentences in the same language (see Nakajima & Takeichi, 2011, for the similar approach with correlation matrix).

Mean RSs in each experimental condition are shown in Figure 3. Note that in this figure “French” and “English” means participants’ speaking language instead of the four sentences’ language. In general, the 350-, 1000- and 2700-Hz conditions led to RSs higher than .57 except French-speaking participants in the peak-equal condition. A *t* test was conducted to examine whether each band led to a significantly higher RS than .57, i.e., with a null hypothesis that the mean RS for each band was .57. Because the between-participants conditions were pooled, the test was conducted four times for the four bands. A significant difference was obtained in 350 Hz [$t(14) = 2.723$, $p = .016$], 1000 Hz [$t(14) = 3.252$, $p = .006$], and 2150 Hz [$t(14) = 2.931$, $p = .011$], but not in 4800 Hz [$t(14) = -.570$, $p = .578$].

An ANOVA according to 2 (participants' language) × 2 (power) × 4 (band) design, with repeated measures on the last factor, revealed that the band effect, $F(3, 33) = 4.708$, $p = .008$, $\eta_p^2 = .300$, as well as the power effect, $F(1, 11) = 8.066$, $p = .016$, $\eta_p^2 = .423$, was significant, while the language effect was not significant, $F(1, 11) = 2.027$, $p = .182$, $\eta_p^2 = .156$. The interaction between the power and the band was significant, $F(3, 33) = 3.023$, $p = .043$, $\eta_p^2 = .216$. The interaction between the language and the power was marginally significant, $F(1, 11) = 3.744$, $p = .079$, $\eta_p^2 = .254$. The remaining interactions were not significant ($F < 1.6$).

Because the interaction between the power and the band was significant, simple main effects were examined. The mean RS changed significantly depending on the frequency band in the mean-equal condition, $F(3, 33) = 7.004$, $p < .001$, $\eta_p^2 = .389$, but not in the peak-equal condition, $F(3, 33) = .728$, $p = .543$, $\eta_p^2 = .062$. In addition, the mean-equal condition led to a significantly higher RS than the peak-equal condition in 350 Hz, $F(1, 44) = 11.181$, $p = .002$, $\eta_p^2 = .203$, and in 1000 Hz, $F(1, 44) = 4.132$, $p = .048$, $\eta_p^2 = .086$.

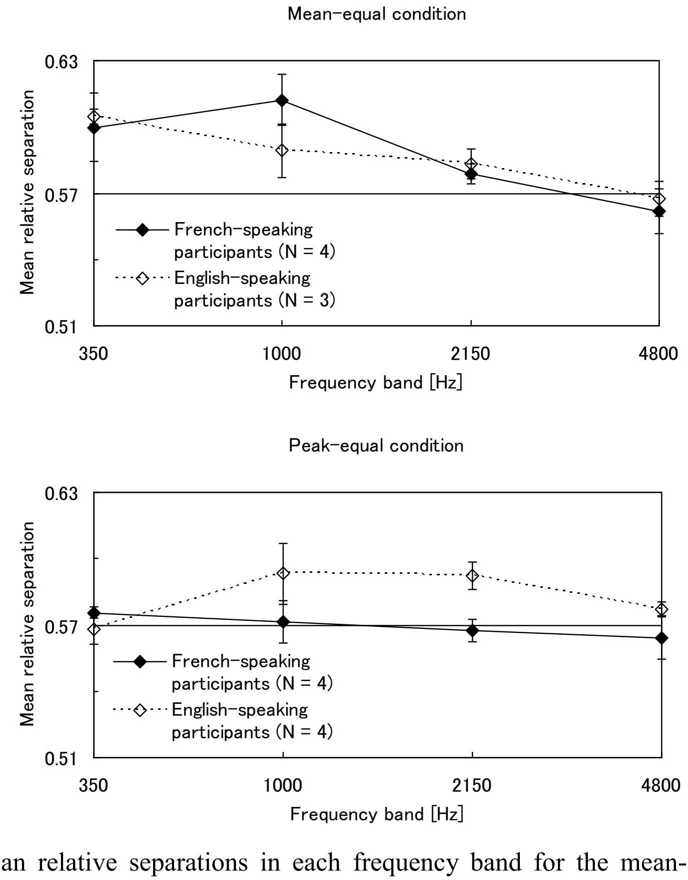


Figure 3. Mean relative separations in each frequency band for the mean-equal condition (upper panel) and for the peak-equal condition (lower panel). Bars are standard error of mean. Note that "French" and "English" means participants' speaking language, instead of the four sentences' language.

Discussion

The present study examined whether French and English rhythms could be discriminated with only temporal changes of intensity included in speech. To avoid presenting any other perceptual cues, e.g., phoneme and pitch, included in speech, the present experiment employed noises that had the same power envelopes as each frequency band of speech had. Because this approach focused on factors determining speech rhythms of French and English, it did not aimed to explain the whole aspects of perceptual difference between these languages. However, the results of the present study indicated that, even without any signs of phoneme and pitch, participants could perceive two power envelopes as more dissimilar rhythms when these envelopes were obtained from sentences in different languages than when obtained from sentences in the same language.

Moreover, the tendency to perceive French and English envelopes as dissimilar was reduced in 4800 Hz compared with the other bands, especially when these envelopes were calibrated at equal RMS levels (in the mean-equal condition). This indicates that French and English rhythms are determined by intensity changes in frequency regions below 2150 Hz. Since this frequency region determines the temporal arrangements of language nucleus, i.e., speech rhythms, of French and English, intensity below 2150 Hz could be utilized for constructing a physical parameter expressing *sonority*, which determines the temporal frames of syllables in phonology. This proposition is consistent with that proposed by Nakajima, Ueda, Fujimaru, Motomura and Ohsaka (2012) who investigated acoustical correlates of sonority in British English with factor analytical approaches.

Acknowledgements

This research was made possible by a research grant awarded to SG by the Natural Sciences and Engineering Council of Canada and grants awarded to YN and to KU by Japan Society of the Promotion of Science.

References

Abelson, R. P. (1954-55). A technique and a model for multi-dimensional attitude scaling. *Public Opinion Quarterly*, *18*, 405-418.

Fant, G., Kruckenberg, A., & Nord, L. (1991). Durational correlates of stress in Swedish, French, and English. *Journal of Phonetics*, *19*, 351-365.

Nakajima, Y., & Takeichi, H. (2011). Human processing of short temporal intervals as revealed by an ERP waveform analysis. *Frontiers in Integrative Neuroscience*, *5*, 2-10.

Nakajima, Y., Ueda, K., Fujimaru, S., Motomura, H., & Ohsaka, Y. (2012). Acoustic correlate of phonological sonority in British English. *Proceedings of the 28th Annual Meeting of the International Society of Psychophysics*. in preparation.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward and understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 756-766.

NTT-AT multilingual speech database 2002 [CD-ROM]. (2002). Tokyo: NTT Advanced Technology, Co.

Patel, A. D. (2008). *Music, Language, and the Brain*. New York: Oxford University Press.

Scharf, B., & Buus, S. (1986). Audition I: Stimulus, physiology, thresholds. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 14-1 to 14-71). New York: John Wiley & Sons.

LISTENING TO SEMANTICALLY ANOMALOUS SENTENCES MASKED BY NOISE AND COMPETING SPEECH IN A SECOND LANGUAGE: A CROSS-LANGUAGE STUDY ON KOREAN-ENGLISH BILINGUALS.

Avivi-Reich[1,2], M., Bae, M.H[1,2]., Kang, Y[3,4].,& Schneider, B.A.[1,2]

1. *Department of Psychology, University of Toronto Mississauga, Canada*
2. *Centre for Research on Biological Communication Systems, University of Toronto Mississauga*
3. *Department of Linguistics, University of Toronto, Canada*
4. *Centre for French and Linguistics, University of Toronto Scarborough*

E-mail: meital.avivi@utoronto.ca

Abstract

Wu et al. (2011) examined masking of speech by speech in Mandarin Chinese listeners. The results of their study showed that Chinese listeners benefitted less from spatial separation than did English listeners when a same-language masker was used; implying that the effectiveness of the spatial separation cue with respect to stream segregation is language dependent. In this study we have asked 24 native Korean speakers to perform a similar task and repeat nonsense sentences in Korean played with one of three types of background noises: 1) steady speech-spectrum noise; 2) same-language speech (Korean); and 3) morphologically similar cross-language speech (Chinese). These conditions were played in 4 different SNRs. The 50% correct corresponding SNR thresholds and the slopes of the psychometric functions were calculated and analyzed. The current results show that when listening to Korean the release from masking seems to be also smaller than when listening in English.

Most daily interactions take place in noisy environments rather than in quiet. This is known to create an acoustic challenge for those who attempt to carry out a conversation in such settings, for some more so than for others. One of the most common complaints made by older adults is related to difficulties they encounter when communicating in noise (CHABA, 1988), however, younger adults may also find these acoustic conditions challenging, especially if they are required to perform while using their second language (L2). A growing number of studies demonstrate the difficulties non-native listeners experience when attempting to perceive and understand acoustically disturbed signal (e.g., Ezzatian, Avivi, & Schneider, 2010; Lu-Feng, 2010; Mattys et al., 2010). Their ability to successfully meet the challenges presented by additional background noise is poorer than that of a native listener even when they have been intensely exposed to the second language from an early age (Ezzatian, Avivi and Schneider, 2010). While more information on the effects that a first language (L1) may have on the speech perception of a second language (L2) is accumulating, the possible effects of intense exposure to a second language (L2) on the speech perception of the first acquired language (L1) remain much less addressed. The question of how an exposure to a second language may affect speech perception performance when listening to the first native language is not less important, since in most cases people do not completely abandon their native language but rather continue to communicate in both. The proportion of people who are using more than one language on a daily basis is increasing because of increases in immigration. In addition, an increasing number of companies and businesses are becoming international in scope, requiring many to communicate in both their first and second language.

From a pragmatic point of view, being able to communicate efficiently in a noisy multi-talker environment is an essential skill. Not being able to efficiently communicate in the presence of other talkers might lead to social isolation and loss of psychological/emotional and other support sources. On a theoretical level, it is important to determine the causes of these difficulties. In nonnative listeners we can consider several possible sources such as a reduced ability to take advantage of context or auditory cues (Ezzatian, Avivi, & Schneider, 2010), limited mastery of the second language's phonetics, slower access to the lexicon, etc. When comparing the results of several studies, which have tried to quantify the contribution of those sources, the complexity of the listening situation seems to be a key variable which must be considered (e.g., Schneider, 2011). In this study we choose to address one acoustic feature which can also be used as a cue when attempting to segregate a target voice from any distracting background sounds, namely spatial separation.

When a target voice is spatially separated from the background noise it allows the listener to take advantage of auditory cues, such as interaural timing differences and SNR differences, to enhance stream segregation. In order to further explore the sources of possible differences in the ability to take advantage of spatial separation, we changed only the apparent location of the target using the precedence effect. If the target voice is played over two loudspeakers located to the left and right of the listener, with the sound on the left loudspeaker slightly lagging behind that on the right, the listener perceives the target voice as emanating from the right. If the masker is played at the same time as the target voice, but with the lag reversed, the masker will be perceived as coming from the left, creating a perceived spatial separation which is achieved without altering the SNRs at each ear. Previous studies, which were done using perceived spatial separation, showed that the ability to use it reduces with age (Murphy et al., 2006) and is somewhat language based (Wu et al., 2010). More specifically, while no difference was found between native and non-native listeners when listening and repeating semantically anomalous sentences in English (Ezzatian, Avivi, & Schneider, 2010), differences were found when monolingual speakers of English and Mandarin Chinese listened in their own language when the target sentence was masked by one of three maskers (same language two-talker speech, cross-language two-talker speech, and speech-spectrum noise; Wu et al., 2010). In the latter study, both groups benefitted equally from spatial separation when the maskers were speech-spectrum noise or cross-language. However, Chinese listeners benefitted less from spatial separation than did English listeners when a same-language masker was used. The results of this cross-language study implied that the effectiveness of the spatial separation cue with respect to stream segregation is language dependent. In order to further explore the effect of L1 and bilingualism on speech perception and comprehension in noise, the current study was done in Korean which is a morpheme-based language similar to Chinese but lacks tonality. More specifically, we wanted to test whether the extent to which spatial separation can be used as a cue may vary based on the type of masker and the age at which an extensive use of the second language has started.

Method

Participants

Twenty four young native Korean-speakers with normal hearing were recruited from University of Toronto, Mississauga. Twelve of the participants arrived to North America before the age of 14 years and 12 arrived at a later age.

Stimuli and Procedure

The grammatically correct but semantically anomalous English Freyman sentences (Helfer, 1997) were translated into Korean and were recorded using a female native Korean speaker. These sentences contain three keywords corresponding to the subject, verb, and object of the simple declarative structure. The order of these components in the Korean version was *subject + object + verb*. The subject and object words of all sentences consisted of two morphemes, whereas the verb did not follow this rule due to structural differences in Korean versus Chinese. These sentences were organized into 24 lists, with 13 sentences in each list, and were played over loudspeakers in a sound-attenuating booth. These sentences were played along with one of three maskers (steady-state speech spectrum noise, two other female talkers producing semantically-anomalous sentences in Chinese, two different female talkers producing semantically-anomalous sentences in Korean). The spatial location of both the target sentences and the masker was varied using the precedence effect. In the co-located condition, the target and the masker were both presented over the right and left loudspeakers with the left loudspeaker lagging by 4ms behind the right, giving rise to the impression that both the target sentences and the masker were emanating from a position to the right of the listener. In the spatially separated condition, while the target sentences on the left loudspeaker were slightly lagging behind that on the right, the masker was played at the same time as the target voice, but with the lag reversed. As a result the target perceived as coming from the right while the masker was perceived as coming from the left, creating a perceived spatial separation. Each combination of masker type and location was played using 4 different signal to noise ratios (SNR): 6, 1, -4, and -9dB, in order to later assess the SNR at which the participants perceives 50% of the target words correctly. Participants were asked to repeat the sentence they heard and each morpheme in each target word was scored as either correct or incorrect.

Results

The initial results of this study show that similar to what was found in Chinese (Wu et al., 2011), Korean native listeners present smaller release from masking due to spatial separation when listening to Korean, than that which was found in English native speakers. In addition, the cross-language two-talker masker (Chinese) interfered with the perception of Korean more than the English cross-language masker did for native Chinese listeners, and more than the Chinese cross-language masker did for native English listeners in Wu et al. (2011). As can be seen in Fig. 1, no significant difference was found in the extent to which perceived separation was used to enhance release from masking between early and late arrivals. This lack of age of arrival main effect implies that the ability to use this acoustic cue is most likely not dependent on individual linguistic experience.

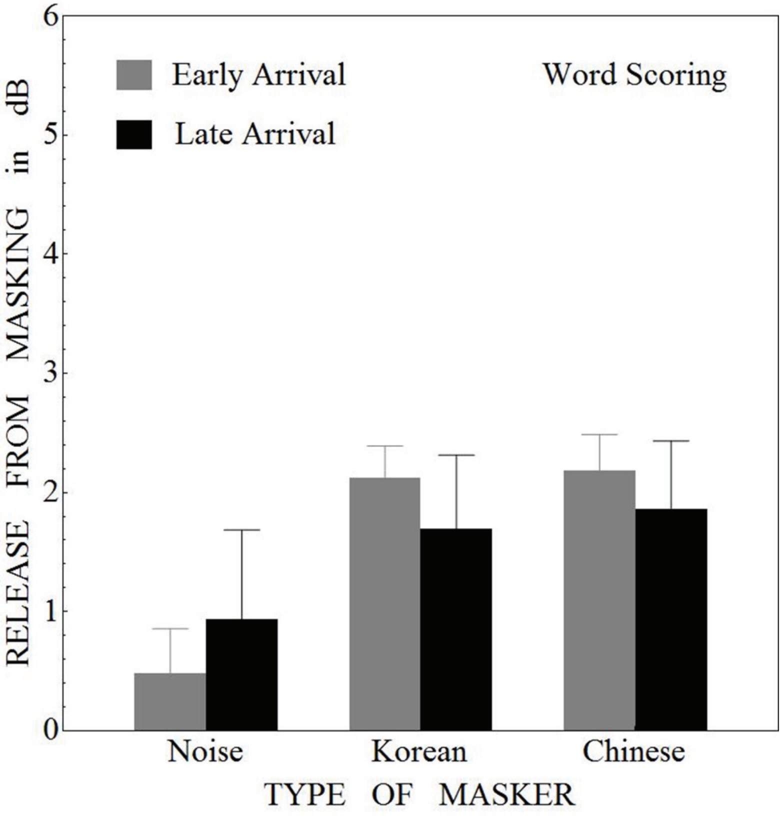


Fig. 1. Mean threshold differences between pairs of maskers for the Conditions in which maskers and targets were co-located versus those where they were, representing release of masking due to spatial separation, separated for Early and Late Arrivals to Canada.

Acknowledgments

This research was supported by a grant from the Natural Science and Engineering Research Council of Canada (RGPIN-9952).

References

CHABA. (1988). Speech understanding and aging. *The Journal of the Acoustical Society of America*, 83, 859 – 895.

Ezzatian, P., Avivi, M. & Schneider, B.A. (2010).Do nonnative listeners benefit as much as native listeners from spatial cues that release speech from masking? *Speech Communication*, *52*, 919-929.

Helfer, K. S., 1997. Auditory and auditory-visual perception of clear and conversational speech. *J. Sp. Lan. Hear. Res*., 40, 432-443.

Lu-Feng, S. (2010). Perception of acoustically degraded sentences in bilingual listeners who differ in age of English acquisition. *Journal of speech, language, and hearing research* : 53 (4), 821 – 835.

Mattys, S.L., Carroll, L.M., Li, C.K.W & Chan, S.L.Y.(2010). Effects of energetic and informational masking on speech segmentation by native and non-native speakers. *Speech Communication*, 52(11), 887 – 899.

Schneider, B.A. (2011). How age affects auditory-cognitive interactions in speech comprehension. *Audiology Research*, 1(1S), 34-39.

Wu, X., Yang, Z., Huang, Y., Chen, J., Li, L., Daneman, M., & Schneider, B. (2011). Cross-language differences in informational masking of speech by speech: English versus Mandarin Chinese. Journal of Speech Language and Hearing Research, 54(6),1506-1524.

Subjective Confidence of Acoustic and Phonemic Representations During Speech Perception

Jordan Richard Schoenherr, John Logan, and Cassandra Larose
psychophysics.lab@gmail.com, john_logan@carleton.ca, clarose@connect.carleton.ca)
Department of Psychology, Carleton University
1125 Colonel By Drive, Ottawa, ON K1S5B6 Canada

Abstract

Acoustic and phonemic information form the basis for speech perception. We used confidence reports to examine the extent to which 1) both representations contributed to identification performance, 2) whether participants had an awareness of acoustic information, and 3) whether confidence reports were influenced by both acoustic and phonemic representations in an identification task. Our results suggest that participants' response were primarily guided by explicit, phonemic information. We also found that an interpolation between typicality ratings and identification functions yielded an excellent fit to the function produced by confidence reports suggesting that confidence processing.

Responses to Two Alternative Forced Choice (2AFC) identification tasks in which listeners are presented stimuli that vary along an acoustic continuum, such as voice-onset time, produce a logistic function defining two groups of stimuli (i.e., /b/ and /p/). These response patterns have been interpreted as a result of listeners utilizing phonemic representations (e.g., Liberman, Harris, Hoffman, & Griffith, 1957), with evidence from discrimination tasks additionally suggesting that participants cannot detect within-category acoustic differences when stimuli are presented at long ISIs (e.g., Pisoni, 1973). The use of multidimensional scaling techniques has also revealed considerable within-category similarity and between-category dissimilarity (Iverson & Kuhl, 1995). Alone, these findings suggest that phonemes might be the only available representations to identify and discriminate stimuli.

Phonemic representations, however, do not appear to be the only source of information available to participants. Using analysis of response latencies, Pisoni (1973; Pisoni & Tash, 1974) proposed a model wherein acoustic properties determine primary decision response selection under certain conditions. Specifically, participants exhibited short response latencies for pairs of stimuli that were identical acoustically (i.e., ΔVOT = 0 ms) as well as those that were highly dissimilar (ΔVOT $\geq$ 40 ms). Longer response latencies were observed for within-category pairs and between-category pairs which had identical acoustic differences (ΔVOT = 20 ms). These differences in response latencies suggested that participants also had access to acoustic information. Further evidence for the availability of acoustic representations also comes from changes in discrimination performance and identification functions. For instance, Pisoni (1973) found that presenting stimuli at short ISIs resulted in increases in accuracy when participants made within-category comparisons (cf. Werker & Logan, 1985). Psychophysical training studies have also found that participants can learn to identify non-native speech sounds (Pisoni et al., 1982). Pisoni et al. (1982) provided participants with three exemplars and feedback after each response. In the 3-category identification, participants' identification functions produced 3 distinct categories. Still other studies using typicality ratings have obtained results suggesting that participants can detect within-category acoustic difference (Miller & Volaitis, 1989).

The existence of acoustic and phonemic representations suggests that multiple representations can be used as evidence for primary decision response selection. Whereas the role of multiple representations has been examined previously in the context of top-down processing in sine-wave speech (e.g., Remez, Rubin, Pisoni, & Carrell, 1981), it is not entirely clear whether participants maintain an awareness of these two sources of information. Methods used that examine calibration of subjective awareness are used here to examine this question.

Quantitative approaches to the assessment of subjective awareness require participants to provide a subjective probability (e.g., 50% represents a guess and 100% represents complete certainty) after performing a task. Underconfidence has generally been obtained in perceptual tasks (e.g., Bjorkman, Juslin, & Winman, 1993) leading some to assert our perceptual system is relatively inaccessible (Dawes, 1980). In comparison, confidence reports obtained for general knowledge questions typically produce overconfidence (e.g., Gigerenzer, Hoffrage, & Kleinbolting, 1991). Challenging these findings, the hard-easy effect (Lichtenstein & Fischhoff, 1977) assumes that participants' subjective bias is determined by task difficulty more generally. For instance, in a line-length discrimination task conducted by Baranski and Petrusic (1994), participants produced overconfident responses for difficult stimulus pairs and underconfident responses for easy stimulus pairs. Similarly, recent studies observed an overconfidence bias in perceptual tasks for stimuli with both perceptual and conceptual properties (e.g., Kvidera, & Koustaal, 2008). In the context of speech perception, we sought to identify overconfident responses due to the presence of acoustic and phonemic representations.

Models of confidence process differ along three dimensions: the locus of confidence processing, the dependency of confidence processing on primary decision processes, and the sources of evidence used to compute confidence. Many early models of confidence assumed a decisional-locus (e.g., Ferrel & McGooey, 1980; Gigerenzer et al., 1991; see also Pleskac & Buseymeyer, 2010) wherein confidence reports are based solely on information used by the primary decision process thereby requiring no additional processing, a post-decisional locus wherein confidence is computed following the primary decision (e.g., Audley, 1960; Vickers & Packer, 1980). A later development was an alterable locus model wherein confidence processing can occur during or after the primary decision depending on speed or accuracy stress and used the total accumulated amount of nondiagnostic evidence to determine certainty (Baranski & Petrusic, 1998). In a study conducted by Baranski and Petrusic (2001) participants were given blocks of trials wherein they were required to simply make a decision or make a decision followed by a post-decisional confidence report. They found that response latencies for the primary decision were significantly longer when confidence was required relative to a no confidence condition indicating an additional set of operations was required to compute confidence. Recent studies have also found that by manipulating the nature of nondiagnostic information available during the primary decision, confidence reports can vary independently of accuracy (Schoenherr, Leth-Steensen, & Petrusic 2010). Applied to phonemic categorization, if acoustic information is available from a perceptual process and phonemic representations are available from the activation of long-term memory representations, then both sources of information should influence confidence reports. Substantial differences in the patterns observed between accuracy and confidence would suggest the existence of acoustic and phonemic representations.

Method

Fifteen and Fifteen listeners from Carleton University students participated in the study for course credit in Experiments 1 and 2, respectively. All participants reported normal hearing and no speech pathologies. Using the paradigm developed by Pisoni and Tash (1974) participants were presented with /b/ and /p/ stimuli that varied along the VOT continuum. Fifteen speech stimuli corresponding to -70 to 70 ms VOT, originally synthesized by Lisker and Abramson (1967), were obtained from the Haskins Laboratories website (HL, 2011). The sounds were originally recorded on reel-to-reel tape and later converted into AIFF format. Stimuli were pre-processed using a DC offset correction to eliminate clicks present in the AIFF versions and then converted into WAV files. Whereas Experiment 1 only used stimuli from the 0 to 60 ms VOT range to replicate Pisoni and Tash (1974), Experiment 2 used the full stimulus range, with stimuli from the -70 ms end corresponding to the prevoiced phoneme category /p^h/ not used phonemically in English, and the remaining stimuli corresponding to the /p/ and /b/ phoneme categories used in English.

Procedure

Trials in the ID task had one or two components depending upon block. In both blocks of trials participants reported whether the stimulus was a /b/ or /p/ using keys labeled B or P on the keyboard ('V' or 'N' key, respectively). For one block participants also rated the confidence they had in their ID responses using a 6-point scale using the 'E' through 'I' keys, with 50% representing a guess and 100% representing certainty. Participants completed a total of 180 trials in each block of the ID task.

Half of the participants performed the ID task first whereas the other half performed the AX task first. Half of the blocks of trials required participants to provide confidence reports whereas the other half only required participants to complete the ID task alone. Presentation of confidence and no confidence blocks was counterbalanced. The experiment required approximately 30 minutes to complete. Stimuli were presented via headphones using PsychoPy software (Peirce, 2007). The procedure was modified in Experiment 2, to include a typicality rating task following the ID task.

Results and Discussion

The results for Experiment 1 and Experiment 2 are provided in Figure 1a and Figure 1b. Identification responses, typicality ratings, and confidence indices were all subjected to repeated measures ANOVAs.

Experiment 1

Proportion Identification. Participants clearly identified two discrete categories for /ba/ and /pa/, respectively, with a category boundary situated between +20 and +30 ms VOT. This pattern replicates the findings obtained by Pisoni and Tash (1974) as well as other studies (e.g., Experiment 1 in McMurray et al., 2003). The proportion of correct ID responses was analyzed for each VOT stimulus and whether a confidence report was provided or not. The only significant finding observed was the location of the stimuli along the VOT continuum, $F(6,84) = 6.394$, $MSE = .019$, $p = .02$, $\eta^2 = .314$. The absence of a main effect or interaction of confidence reports is important as it suggests that the addition of confidence reports did not significantly affect ID performance thereby permitting a straightforward interpretation of the remaining results.

Confidence Reports. Figure 1 also demonstrates the effect of confidence measures. Like ID accuracy, we found that subjective confidence varied along the VOT continuum, $F(1,14) = 6.55$, $MSE = 44.11$, $p = .008$, $\eta^2 = .319$. Pairwise comparisons revealed that this effect arose from the difference in confidence between stimuli located at 20 and 30 ms VOT ($p = .035$), which corresponds to the stimuli adjacent to the category boundary. An analysis of subjective calibration revealed only a marginally significant difference across the VOT continuum, $F(6,84) = 3.401$, $MSE = .013$, $p = .085$, $\eta^2 = .195$. This suggests that the greatest difference between subjective awareness and performance occurs for the 20 ms VOT stimulus. Our comparison of over/underconfidence bias did not reveal any significant effects, $F(6,84) = 1.948$, $MSE = .035$, $p = .183$, $\eta^2 = .122$. Together, these findings suggest that participants are only explicitly aware of the phonemic representation.

Figure 1a.

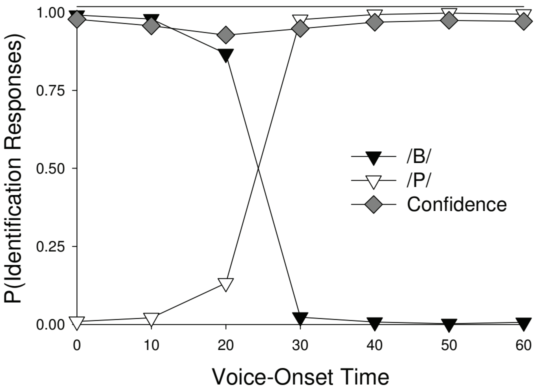


Figure 1b.

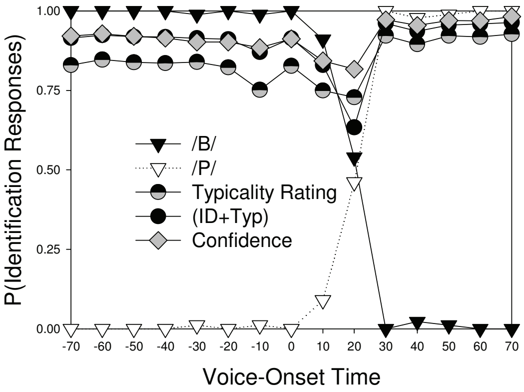


Figure 1a. Mean identification functions, response times for confidence (unfilled circles) and no confidence (filled circles) conditions and mean confidence across VOT continuum. The identification function uses performance in confidence condition to allow comparison with mean confidence. Figure 1b. Mean identification functions, typicality ratings, confidence reports, and interpolation line for Experiment 2.

Experiment 2

Proportion Identification. Replicating the general results of Experiment 1, the location of the stimuli along the VOT continuum significantly affected identification performance, $F(14,112) = 9.149$, $MSE = .124$, $p = .005$, $\eta^2 = .533$. Figure 1b demonstrates, participants had a sharp category boundary between stimuli for the /b/ and /p/ categories. A noticeable difference was evident in the location of the boundary. Whereas in Experiment 1 the boundary was located between VOT 20 ms and 30 ms, a shift such that the boundary was now located at VOT 20 ms with a resulting decrement in performance for VOT 10 ms stimuli. We can take these results as indicative of range effects. In general, these findings permit a straightforward interpretation of the remaining results.

Confidence Reports. Figure 1 also demonstrates the effect of confidence measures. Like ID accuracy, we found that subjective confidence varied along the VOT continuum, $F(1,14) = 6.55$, $MSE = 44.11$, $p = .008$, $\eta^2 = .319$. Relative to Experiment 1, we did observe greater underconfidence in the negative portion of the VOT continuum.

Typicality Task. The analysis of typicality ratings also obtained a significant result of stimulus location along the VOT continuum, $F(14,112) = 5.820$, $MSE = .3.295$, $p = .009$, $\eta^2 = .421$. Unlike accuracy, but like mean confidence, typicality ratings appeared to be more responsive to the acoustic properties of the stimuli. Participants considered stimuli in the /b/ and /p^h/ range as less typical then stimuli in the /p/ range even though they exhibited equal accuracy. Moreover, within-category ratings exhibited more graded responses.

Interpolated Function. The similarities in patterns observed in confidence and typicality suggested a potential relationship between these two functions. As Figure 1b suggests, mean confidence ratings are situated between accuracy in the identification task and typicality ratings in the typicality task. Pearson's correlations revealed the strongest relationship between confidence and typicality ratings, $r^2 = .960$, $p < .001$. The correlations between identification responses and mean confidence was also significant, $r^2 = .446$, $p = .007$, although the correlation between identification and typicality was only marginally significant, $r^2 = .261$, $p = .051$. These findings suggest that confidence is associated with both identification accuracy and typicality ratings but that identification accuracy and typicality ratings are only weakly related.

In order to examine the relationship between accuracy, typicality, and confidence ratings we converted typicality to a proportion, summed it with proportion correct, and produced an interpolated function. A paired-samples t-test revealed that the mean confidence function and the interpolated function did not significantly differ from one another, $t(14) = .309$, $p = .762$. This suggests that confidence reports were closely associated with information from both identification accuracy (associated with phonemic representations) and typicality ratings (associated with acoustic information). All other paired-sample t-tests were significant (all *t*s > 3.283, *p*s < .005) indicating that different sources of information contributed to response selection for each dependent measure.

General Discussion

In general, we obtained results consistent with previous studies of speech perception and confidence processing. A 2AFC identification task using stimuli from the prevoiced-unvoiced VOT continuum produced responses suggesting two phoneme categories were involved in response selection. When that range was restricted to voiced and unvoiced stimuli (Experiment 1), participants only exhibited overconfidence around the category boundary. This finding supports the claim that overconfidence can be obtained in perceptual tasks (e.g., Baranski & Petrusic, 1994). Moreover, overconfidence in this context also suggests that a phonemic representation is used to identify stimuli. Extending the range to prevoiced stimuli (Experiment 2) resulted in underconfidence, suggesting that acoustic properties of these stimuli were in fact available to participants even if they were not used in primary decision response selection. Thus, the results of the present study support the availability of two kinds of stimulus representations - acoustic and phonemic – that influence response selection (cf. Pisoni, 1973; Pisoni & Tash, 1974).

An important result obtained in the present study was the relationship between typicality ratings and confidence reports. Typicality ratings were provided by participants because previous studies obtained results suggesting that these ratings were affected by acoustic information (e.g., Miller & Volatis, 1989). We initially assumed that confidence judgments might be influenced by acoustic information, producing more continuous responding. In Experiment 1, we observed a trend suggesting that confidence was only affected by phonemic information. In contrast, the results of Experiment 2 acoustic properties reduced subjective confidence and typicality ratings in the prevoiced portion of the continuum. When an interpolated function was obtained for identification accuracy in an identification task and typicality ratings in a typicality task, it yielded an excellent fit to the confidence function in the identification task. In the absence of another explanation, it seems reasonable to suggest that confidence reports are the product of both acoustic and phonemic representations. These results could be taken as support for a doubt-scaling model of confidence processing (Baranski & Petrusic, 1998). In the present task, participants were presented with /p^h/ stimuli which that are outside the range of their native phoneme categories (although Lisker & Abramson, 1964, report that stops produced by some English listeners incorporate prevoicing, it is not used phonemically in English). Namely, although the identification task only requires the use of phonemic representations that are activated in long-term memory as a result of accumulated acoustic evidence, that acoustic information in the prevoiced region increases the amount of uncertainty as to category membership. Thus, while primary decision accuracy has reached a performance asymptote, subjective confidence is reduced due to the availability of this evidence.

References

Abramson, A., & Lisker, L. (1965). Voice onset time in stop consonants: Acoustic analysis and synthesis. Proceedings of the Fifth International Congress on Acoustics, Liege, A51.

Baranski, J. V., & Petrusic, W. M. (1994). The calibration and resolution of confidence in perceptual judgements. Perception & Psychophysics, *55*, 412-428.

Baranski, J. V., & Petrusic, W. M. (1998). Probing the locus of confidence judgments: Experiments on the time to determine confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 929-945.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *98*, 506-528.

Haskins Laboratories (2011). Abramson/Lisker VOT Stimuli. Retrieved 01/12/2011. From http://www.haskins.yale.edu/featured/demo-liskabram/index.html /.

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America, 97,* 553–562.

Kvidera, S., & Koustaal, W. (2008). Confidence and decision type under matched stimulus conditions: overconfidence in perceptual but not conceptual *Decisions. Journal of Behavioral Decision Making*, *21*, 253–281.

Liberman, A.M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*, 358-368.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how, much they know? *Organizational Behavior and Human Performance*, 20, 159-183.

Lisker, L, & Abramson, A. S. (1967). The voicing dimension: Some experiments in comparative phonetics. *Proceedings of the 6th International Congress of Phonetic Sciences.* Prague: Academia.

Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, *46*, 505-512.

Peirce, J. W. (2007) PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*, 8-13.

Pisoni, D. B. (1973). Auditory and phonetic memory codes inthe discrimination of consonants and vowels. *Perception & Psychophysics, 13,* 253-260.

Pisoni, D. B., Aslin, R. N., Percy, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 297-314.

Pisoni, D. B., & Tash, J. B. (1974) Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics, 15,* 285-290.

Remez, R.E., Rubin, P.E., Pisoni, D.B., Carrell, T.D. (1981). Speech perception without traditional speech cues. *Science, 212*, 947-949.

Vickers, D., & Packer, J. S. (1982). Effects of alternating set for speed or accuracy on response time, accuracy, and confidence in a unidimensional discrimination task. *Acta Psychologica*, 50, 179-197.

Werker, J. F. & Logan, J. S. (1985). Cross-language evidence for three-factors in speech perception. *Perception & Psychophysics*, *37*. 35-44.

IMPLIED MOVEMENT PERCEPTION IN DIFFERENT STATIC ARTWORKS AFFECTS SUBJECTIVE TIME

José Lino Oliveira Bueno and Francisco Carlos Nather
Center for Experimental Aesthetics, University of São Paulo, School of Philosophy, Sciences and Letters of Ribeirão Preto, São Paulo, Brazil
jldobuen@ffclrp.usp.br and f.nather@hotmail.com

Abstract

Several experiments from our laboratory are showing that paintings and sculptures with more implied movement were estimated longer than those artworks with lesser movement. Furthermore, presentations of short and long term duration of abstract and figurative artworks differently distorted the perception of time. These time distortions have been related to changes in levels of arousal associated to embodiment mechanisms and to different cognitive processes as memory and attention. Aspects of different movement representation in static artworks of different artistic movements (for example impressionism, cubism and op art) are associated to different duration exposures on subjective perception of time. Eye tracking and force platform procedures confirm the relationship between implied movement perception and subjective time. The aesthetic experience of art is not limited to vision and cognitive components of the observer, but involves different aspects of work-spectator relation as the representation of movement, emotions and mechanisms of embodiment.

1.

It is well known the relationship between perception of objects in motion and the subjective experience of time. However, static objects can, surprisingly, induce the perception of motion. Will this perception of implied motion affect also the subjective time?

Several experiments from our laboratory showed that pictures of paintings and sculptures exposed to the participants for the same duration but with more implied movement were estimated longer than those with lesser movement. The temporal distortions were similar to those obtained with visual stimuli in real movement. In a first study, Nather and Bueno (2006) showed that the picture of an artwork sculpture of a doll with less suggestion of implied movement was reproduced as shorter duration than that one with greater suggestion of movement. Nather and Bueno (2008), in sequence to this study, developed the Body Movement Ranking Scale (BMRS), which allows identification of the degree of movement implied by body position in still images, using photographs of the sculptures of the impressionist Edgar Degas ballerinas. The results obtained from the use of this scale showed that dancers standing on steps or with small movements (1.5 and 3.0 points) were reproduced as with shorter duration than ballerinas dancing (4.5 points); standing on steps or dancing ballerinas were estimated shorter than those that represent major steps of classical ballet (6.0 points). Consistent with these data, the dancers with 1.5 and 3.0 (points) were underestimated in relation to the actual duration of stimulus presentation, but those with 6.0 points were overestimated.

The use of BMRS scale showed that the visual processing of motion induced by static images can modulate the estimation of experienced time (Nather & Bueno, 2011).

Furthermore, the data can be interpreted as a processing similar to temporal bisection, which depends on the distances among the values of the series of BMRS stimuli presented to the participants: individuals tend to estimate as having similar duration images with closest movement intensities; for example, there are no differences between the time estimations of 4.5- and 6.0-point stimuli, but both are different from the 1.5-point stimulus; at the same time, there are no differences between the time estimations of 1.5- and 3.0-point stimuli, but both estimations are different of the 6.0-point stimulus. Furthermore, a stimulus can be estimated with different durations depending on the relationship between its movement score and different scores of the others with which it is presented.

2.

Distortions in subjective time on static images, which may be occurring due to induced or represented movements, have found support in other experiments that evaluated aspects of visual perception of motion: eye movements and body oscillations.

Nather, Bueno and Bigand (2009) examined the eye movements to describe where and how often body parts in different positions were observed in a study of implicit movement and subjective time. Stimuli were photographic images of 3 Degas ballerinas sculptures ordered by BMRS in geometric progression. Images were presented to participants whose tasks were to observe the images and estimate the presentation time duration. Data analysis showed that the same duration static images with 1.5 and 3.0-point stimuli were estimated as shorter than 6.0-point; 1.5 and 3.0-point were underestimated; 6.0-point were overestimated regarding real time. Most attention to arms and legs was observed in the sculpture with more movement and more attention to head and trunk to sculptures with less represented movement. These data suggest that visual perception of movement modulates the pattern of eye movements which are related to the time distortions.

Nather, Bueno, Abreu and Gomes (2010) considered that the observation of body movements in static images generates the experience of movement that can induce real movements in the observer. This study examined whether these real movements were related to the intensity of the observed movement. Stimuli were photographic images of 2 dancer sculptures (BMRS): 1.5- and 6.0-point stimuli. Images were presented at a random order to the participants positioned on a force platform, whose tasks were to observe each image and estimate its presentation time duration. Participants moved more when they observed the dancing ballerina (6.0-point). Also, 6.0-point stimulus was overestimated. This result showed that images of body movements internally generate unconscious body oscillations suggesting that different processes are involved in the subjective time distortions.

Thus, distortions in the subjective time perception of images with induced movement may be attributed to the activation of specific neurons related to visual perception of movement.

3.

These data employing segments of artworks, presented with longer durations than that more usual in traditional studies of perception of time, raise special issues. Could these data on the aesthetic appreciation and perception of movement be limited to the parameters of this experiment? Another characteristic of these events could be identified as important in the subjective perception of time. Among the variables present in these studies, in addition to the issue above discussed, the experimental procedures and duration of the stimuli are important components to be considered.

Using the same method of previous cited works (reproduction method), Nather & Bueno (2012c) presented the sculptures of Edgar Degas for 9, 18, 27 or 45 seconds (G9, G18, G27 and G45 groups, respectively) and the stimuli were randomly presented in arithmetical (1.5-, 3.0- and 4.5-point) or geometrical (1.5-, 3.0- and 6.0-point stimuli) progressions. Data analysis showed that time was not distorted in G9, G18 and G45 groups, except: 6.0-point stimulus was overestimated in geometrical (G9) and 1.5-point was underestimated in arithmetical (G45) progressions. However, time distortions in G27 group were modulated by different implied movement intensities as was observed in previous works that used 36 s of image exposure. These results show that different processes evolving the visual perception of movement in static images are also associated to the different exposure duration. That is, the duration of the stimulus should be considered for studies of subjective time and induced movement perception of artworks.

The time estimation method may also affect the subjective perception of time under the experimental conditions described above. Nather, Bueno, Bigand and Droit-Volet (2011),using a temporal bisection task with two ranges of standard durations (0.4-1.6 s and 2-8 s), investigated whether the perception of presentation durations of 1.5- and 6.0-point stimuli was distorted as function of the embodied movement that originally produced these stimuli of BMRS. The participants had to judge whether the presentation duration of each of the pictures was more similar to the short or to the long standard duration. The results showed that the duration was judged longer for the stimulus requiring more movement than for the stimulus requiring less movement as was observed in previous studies using long durations (36 s). The authors related these data with an arousal effect of limited duration on the speed of the internal clock system, once low-arousal body posture was judged to require no movement and the other with a high-arousal body posture was judged to require considerable movement.

Another method to examine the time perception is the exploration time of an event. Nather and Bueno (2012b) allowed the participants to observe the 1.5-, 3.0-, 4.5-, and 6.0-point stimuli for any length of time (exploration time) and, immediately after each image was observed by the participants they recorded the duration as perceived. The results of temporal ratio (exploration time/time estimation) showed that exploration time of images also affected perception of time, i.e., the subjective time for sculptures representing implied movement were overestimated. Together with data obtained using fixed time of 36 s, the authors concluded that long durations of exposure involve more complex cognitive mediation, as attention and memory of the events.

4.

The aesthetic experience, particularly in the appreciation of works of visual art, assumes exposure to stimuli or longer duration pictures than those often used in studies of visual perception and subjective time. The intrinsic properties and characteristics of the artistic appreciation must, therefore, affect the perception of induced movement and the subjective experience that accompanies it. In this sense, it is important to consider the properties of the image used as its pictorial composition.

Nather (2006) and Nather and Bueno (2012a) argue that the scene of a ballet painting by Edgar Degas may implicitly contain narratives which involve different time durations. In this sense, focusing parts of a scene, for example the body of a ballerina in a scene of a ballet choreography, reveals a time that may be related to perceived implicit movement between the body parts. Expressive qualities of movement are connected to what we know about their meaning. The photograph of a dancer gives the observer dynamic properties, because the position of the body in a pose of ballet is perceived as a deviation from the normal. More than reference points for the eyes, the parts of the body not only direct the eye, but also show what the body is doing: in *Dancer Posing at a Photographer's studio*, the dancer's arms are not directed at the top, but raised as in a specific classical ballet step suggesting that some properties and functions of the body constitute an inseparable part of its visible character. One might suspect intentionality on the part of the artist to paint or sculpt movement, representing it in the intensity that he specifically wanted. This would explain why the time may be distorted differentially when participants see the sculptures with different movement intensities, as consequence of the asymmetries generated by the specific characteristics of the diverse parts of the body.

Can the movement implicit in more abstract artworks (paintings) affect the perception of time in the same way that when individuals observed figurative paintings and sculptures of Edgar Degas?

Using 20 abstract paintings implying different types of movement that were exposed by 3 s in random order sequences, Nather, Fernandes and Bueno (2012) showed that cubist paintings (Georges Braque, Juan Gris and Pablo Picasso) representing human figures were differently perceived: the painting with greater arousal and more implied movement was estimated longer than the paintings with lesser arousal and movement. These results are in agreement with those that used figurative human bodies exposed by short and long-term durations which were explained by embodiment mechanisms. The perception of time can be affected by different pictorial characteristics of artworks. Also, Nather, Mecca and Bueno (2012) using two optical paintings of Bridget Riley exposed for 9 or 36 s (G9 and G36 groups) showed no time distortions in G9 group. In the G36 group the paintings were differently perceived: 6.0-point was estimated longer than 2.0-point. Also, the exhibition time of the 2.0-point painting was underestimated, compared with the real time. These results show that also optical illusion of movement in static images caused time distortions related to long duration of exposures.

Artworks employ different experiences of movement perception. In this way, short and long term duration stimuli involve different types of implicit movement and different processing of subjective time, originated from the interaction of various variables or factors, including duration, type of movement, context of the art observation, etc.

The literature of movement perception and, more specifically, the perception of movement in artwork static images has shown that specific brain areas respond differentially to the perception of movement, as in abstract works, as in figurative images of human bodies. More than that, the representation of movements in different artworks has been associated with mechanisms of embodiment that point to the fact that the movement represented in a work of art would be experienced in the body of the observer.

The data presented in this paper enable the establishment of relations between embodiment mechanisms and the subjective time perception. Since the movement occurs at the intersection of space and time, our studies allow the examination of the time representation implicit in the artworks. Works of art are visual objects related to the real life which can represent the movement and the time in a diversity of ways.

In this sense, the aesthetic encounter and the relationships that are established between the observer and the artwork are added when different implicit movements, different durations of observation and different ways of estimation of the observed time are studied in different experimental contexts.

5.

The temporal distortions and illusions following the visual processing can be explained by different perceptual and cognitive models of subjective time. Works of art representing more implicit movement generate greater arousal in individuals, causing temporal over estimations, due to the acceleration of a neural marker. Because they are more complex, they require more processing and memory space, being overestimated. The works with more movement should induce greater directed attention by generating the expectation of the movement. In the case of figurative works, for example, the "paralyzed" image of a dancer in a great ballet step provides clues about how the movement would occur in space, generating the expected time needed for completion.

Our studies show that the aesthetic experience of works of visual art is not limited to vision and cognitive components of the observer, but involves different aspects of work-spectator relationship: collative properties of the artworks as the representation of movement, emotions and embodiment mechanisms.

The artistic appreciation is involved in temporal processing of stimuli and the longer duration of these conditions can be a condition for aesthetic episode, mobilizing not only basic psychological processes such as attention and memory, but also mechanisms of embodiment.

Acknowledgements

Francisco C. Nather received a Postdoctoral Scholarship and a Young Investigators Awards support both from São Paulo Research Foundation (FAPESP). José L. O. Bueno received a Research Grant from the National Counsel for Technological and Scientific Development.

References

Nather, F. C. (2006). *Percepção de movimento e tempo subjetivo nas artes visuais*. PH. D. Th.., University of São Paulo.

Nather, F. C., & Bueno, J. L. O. (2006). Efeito de imagens estáticas com diferentes representações de movimento sobre a percepção subjetiva de tempo. *Psicologia: Reflexão & Crítica*, *19*, 217-224.

Nather, F. C., & Bueno, J. L. O. (2008). Movement ranking scale of human body static images for subjective timing estimation. In *Proceedings of the Annual Meeting of the Society for Psychophysics*, *24*, 185-190.

Nather, F. C., Bueno, J. L. O., & Bigand, E. (2009). Time estimation and eye-tracking movement in human body static images. In *Proceedings of Annual Meeting of the Society for Psychophysics*, *25*, 399-404.

Nather, F. C., & Bueno, J. L. O. (2011). Static images with different induced intensities of human body movements affect subjective time. *Perceptual and Motor Skills*, *113*, 157-170.

Nather, F. C. & Bueno, J. L. O. (2012a). Timing perception in paintings and sculptures of Edgar Degas. *Kronoscope: The Journal for the Study of Tim*e, *12*, 16-30.

Nather, F. C. & Bueno, J. L. O. (2012b). Exploration time of static images implying different body movements causes time distortions. *Perceptual and Motor Skills*, *115*, 105-110.

Nather, F. C. & Bueno, J. L. O. (2012c). The presentation of long term duration of body movement in impressionist artworks differently distorts the perception of time. In *Proceedings of the Annual Meeting of the International Society for Psychophysics*, *28*.

Nather, F. C., Fernandes, P. A. M., & Bueno, J. L. O. (2012). Timing perception is affected by cubist paintings representing human figures. In *Proceedings of the Annual Meeting of the International Society for Psychophysics*, *28*.

Nather, F. C., Mecca, F. F., & Bueno, J. L. O. (2012). Motion illusions caused by paintings of Op Art distort the perception of time. In *Proceedings of the Annual Meeting of the International Society for Psychophysics*, *28*.

Nather, F. C., Bueno, J. L. O., Abreu, D. C. C., & Gomes, M. M. (2010). Body movements and timing estimation related to visual observation of different images representing distinct body positions. In *Proceedings of Annual Meeting of the Society for Psychophysics*, *26*, 433-438.

Nather, F. C., Bueno, J. L. O., Bigand, E., & Droit-Volet, S. (2011). Time changes with the embodiment of another's body posture. *PLoS ONE*, *6*, 19818.

A NEW PERSPECTIVE ON VISUAL WORD PROCESSING EFFICIENCY

James T. Townsend
Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana, USA
Joseph W. Houpt
Department of Psychology, Wright State University, Dayton, Ohio, USA

Abstract

As a fundamental part of our daily lives, visual word processing has received much attention in the psychological literature. Despite the well established perceptual advantages of word and pseudoword context using accuracy, a comparable effect using response times has been elusive. Some researchers continue to question whether the advantage due to word context is perceptual. We use the capacity coefficient, a well established, response time based measure of efficiency to provide evidence of word processing as a particularly efficient perceptual process to complement those results from the accuracy domain.

Due to the relative ease with which most adults read, it is reasonable to assume that word perception is an efficient process. This is further supported by the intuition that with more experience with a process we become more efficient and we are quite experienced with the written word. Often, the efficiency is measured using single letter perception as a base line. When word context offers an advantage in the accuracy or processing time of perceiving a letter, this supports the claim that word perception is efficient.

From the early days of experimental psychology, researchers have been interested in the value of a word context for perceiving letters. In one study, letters were displayed sequentially to participants at faster and faster rates until they could no longer correctly identify the letters. They found that participants maintained accuracy with shorter durations when the letters were presented as part of a word compared with random letter sequences (Cattell, 1886). One problem with studies of this nature is that they do not control for the constraint on possible letters that a word context puts on the possible letters. Hence it is not clear from those early results whether the advantage is a perceptual advantage or a decisional advantage. In the late 1960's an alternative task was designed to eliminate the decisional advantage of word context so as to examine the perceptual effects. In this task a letter or word was tachistoscopically displayed to a participant. They then chose from two possible choices, one of which was correct. In the letter condition, the choices were letters. In the word condition, both choices were words that differed in only a single letter. Since both alternatives were words, the word context was no longer informative as to the identity of the letter. Participants were still more accurate at perceiving letters in the word condition than the letter condition (Reicher, 1969). Furthermore, they found that participants are also more accurate with word contexts than random letter sequence contexts (the word superiority effect). An efficiency gain of context over letters alone is not unique to words though. If a sequence of letters conformed to the pronunciation rules of English, strings referred to as pseudowords, then participants were again more accurate than letters alone (the pseudoword superiority effect, e.g., McClelland & Johnston, 1977).

Despite the robustness of the word and pseudoword superiority effects, a comparable effect using response times (and controlling for decisional information due to context) has been elusive. This may be in part explained by the possibility that people will read an entire word even if the task does not require it. Indeed, this has been put forth as further evidence that word perception is special (LaBerge & Samuels, 1974). One of the goals of this paper is to demonstrate a response time based word superiority effect, and possibly a pseudoword superiority effect as well. In the next section we describe the capacity coefficient, a response time based measure of efficiency. We propose that this measure, along with a task that controls for both the available information and possibly mandatory word reading, provides evidence of word processing as a particularly efficient process to complement the accuracy results.

The Capacity Coefficient

The capacity coefficient, $C(t)$ is an established response time based measure of the effect of increased load on processing efficiency (Townsend & Nozawa, 1995; Townsend & Wenger, 2004). Specifically, $C(t)$ is a measure of the change in processing rates as the task requires attention to more targets, or possibly more dimensions of a single target. The basic idea of the measure is to compare response times when reading the full string to the times that would be predicted if each character took the same amount of time, whether or not it was in a string.

The capacity function for an exhaustive task is defined using the natural log of the cumulative distribution function, $K(t) = \ln F(t); F(t) = \Pr\{\mathrm{RT} \leq \mathrm{t}\}$, and is similar to the cumulative hazard function used in survival analysis. If K_{c1} is the cumulative hazard for the first character response times, K_{c2} is the cumulative hazard for the second character, etc., and K_S is the cumulative hazard for the string condition, the capacity coefficient is given by $C(t) = \left[\sum_{i=1}^{4} K_{c_i}\right] / K_S$.

This formulation is based on the predictions of the unlimited capacity, independent, parallel (UCIP) model. The assumptions of the UCIP model are sufficient conditions for there to be no change in the rates of processing with increased load. If these assumptions hold then the relationship between the processing times of the string to the processing times of the individual characters is as follows:

$$\Pr\{RT_S \leq t\} = \Pr\{RT_{c_1} \leq t\}\Pr\{RT_{c_2} \leq t\}\Pr\{RT_{c_3} \leq t\}\Pr\{RT_{c_4} \leq t\}$$

By taking the natural log of both sides of this equation, then dividing by the left hand side, we see that the UCIP model predicts $C(t) = 1$ for all $t \geq 0$. This gives us a baseline for comparison. If a person performs better than the baseline model, $C(t) > 1$, their performance is referred to as super-capacity. There are multiple ways performance could be super-capacity. For example, if there is facilitation between the characters, or in more extreme cases if the information from the characters is accumulated together toward a single decision (Townsend & Wenger, 2004). Performance worse than the baseline model, $C(t) < 1$, is limited-capacity. In contrast to the case of super-capacity, inhibition between characters could result in limited-capacity. When performance is about the same as the baseline model, $C(t) \approx 1$, then we refer to it as unlimited capacity.

The capacity coefficient measures processing efficiency in isolation by comparing the capacity coefficient to predicted values of unlimited capacity, independent, parallel models. Thus, this measure also allows us to compare the efficiency of a variety of processes despite any possible differences in difficulty due to component processes. In particular, we are able to draw conclusions about the efficiency of word processing relative to pseudoword, non-word, upside-down non-word, and unfamiliar character string processing.

Method

To compare perceptual efficiency across words, pseudowords, non-words, upside-down words and unfamiliar characters, our task must eliminate the extra information available given a word

	Target	Distractors				Single Character							
Word	care	bare	cure	cave	card	c	b	a	u	r	v	e	d
Pseudoword	lerb	nerb	larb	lemb	lerf	l	n	e	a	r	m	b	f
Non-Word	rlkf	vlkf	rtkf	rlhf	rljk	r	v	l	t	k	h	f	k
Upside-down	ɟʞlɹ	ɟʞlʌ	ɟʞʇɹ	ɟɥlɹ	ʞſ̣lɹ	ɹ	ʌ	l	ʇ	ʞ	ɥ	ɟ	ʞ
Katakana	サイクオ	ヘイクオ	サナクオ	サイフオ	サイクノ	サ	ヘ	イ	ナ	ク	フ	オ	ノ

Table 1: Stimuli used for capacity analysis.

context. The possibility that words are exhaustively processed automatically may lead to a disadvantage for words on response time measures. To address these issues, we used a task which forces exhaustive processing of the characters in a string. This experiment consists of two components. We measured the participants' response times to correctly identifying the target string. To ensure that participants identified targets using the entire string and not any subset, we included a distractor of a string with a single character in each position different from the string. For example, if the target is "care" then "bare," "cure," "cave" and "card" were used as distractors (see Table 1). The participants also distinguished between each letters in isolation. Whereas in the exhaustive case the participant needed to distinguish between "bare" and "care," participants distinguished between "b" and "c" in this condition.

Participants were recruited from the Indiana University population. Eight females and two males participated in this study, all of whom were native English speakers and reported that they did not read or speak Japanese. Their ages ranged from 19-34. All participants reported having normal or corrected to normal vision, no difficulty reading English, and no prior diagnoses of a reading disorder.

Table 1 summarizes the stimuli used for both the single character and exhaustive trials for each type. There were five types of stimuli used: words, pronounceable non-words (pseudowords), unpronounceable non-words, upside-down unpronounceable non-words, and strings of Katakana characters. All strings used were four characters long. Words were chosen so that the frequency of the target was roughly equal to the average frequency of the distractors. Pseudowords were taken from the ARC Nonword Database (Rastle, Harrington, & Coltheart, 2002). Strings and characters were presented in black Courier font on a gray background.

Participants were paid $8 per session, and received a $20 bonus upon completion of all 10 sessions. Each session lasted between 45 and 60 minutes and was dedicated to one of the five types of stimuli (e.g., word, pseudoword, . . .), so there were two sessions of each type. At the beginning of each session, we read the participant the general instructions for the task while those instructions were presented on the screen. The instructions encouraged participants to respond as quickly as possible while maintaining a high level of accuracy. Each session was divided into five blocks, one block of string stimuli and a block for each of the corresponding single character stimuli.

Each block began with a screen depicting the button corresponding to each of the categories. Participants had 40 practice trials, 20 of each category. Next participants were given 240 trials divided evenly between the two categories, the first 40 of which were not used in the analysis. Each trial began with a 30 ms presentation of a fixation cross. After a random delay (300-600 ms), the stimulus was presented for 80 ms. Participants had a maximum of 2500 ms to respond. If the participant responded correctly, the next trial started after a 400 ms delay. If the participant responded incorrectly, a tone was played during the 400 ms delay. The session order was counterbalanced among the participants so that participants completed the different

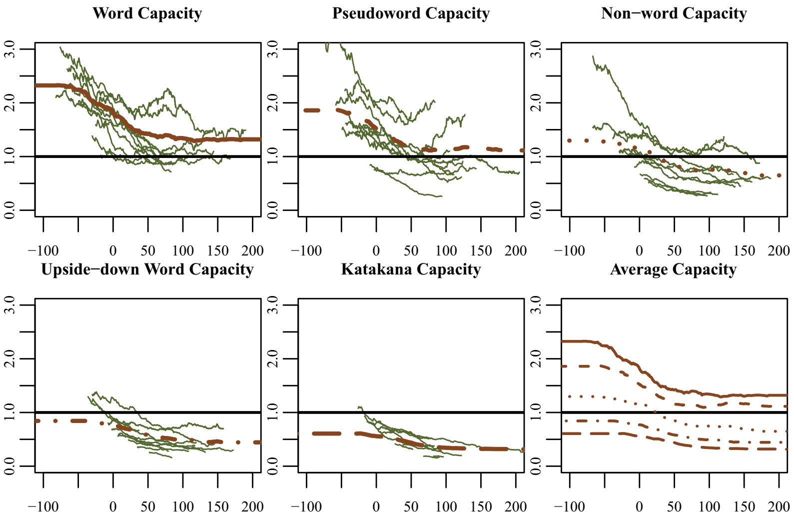


Figure 1: Capacity coefficient functions for each version of the task. The thin lines represent individual participants' data and the thick line represents th average across functions.

types on different days and in different orders.

Results

Individual capacity coefficients are shown in Figure 1. Z-scores for individual and group data, using the statistic in Houpt and Townsend (2012) are shown in Table 2. Each Z-score indicates a test of the null-hypothesis that a participant performs equally to a UCIP model. Significance values are based on a two-sided test. Nearly all participants are significantly different from UCIP, usually better in the word and pseudoword conditions and worse in the other conditions.

Using repeated measures ANOVA, we found a significant effect of condition on capacity ($F(4,36) = 22.64, p < 0.05, \eta^2_G = 0.58$). For post-hoc analyses, we used the z-scores resulting from the mean difference between subjects' capacity z-scores in each pair of conditions. Word capacity was significantly higher than pseudoword capacity ($z = 7.27, p < 0.0025$), random letter capacity ($z = 22.9, p < 0.0025$), upside-down capacity ($z = 36.7, p < 0.0025$), and Katakana capacity ($z = 45.9, p < 0.0025$). Pseudoword capacity was significantly higher than random letter capacity ($z = 15.6, p < 0.0025$), upside-down capacity ($z = 29.4, p < 0.0025$), and Katakana capacity ($z = 38.6, p < 0.0025$). Random letter capacity was higher than upside-down capacity ($z = 13.8, p < 0.0025$), and Katakana capacity ($z = 22.9, p < 0.0025$). Upside-down capacity was significantly higher than Katakana capacity ($z = 9.19, p < 0.0025$).

Discussion

Due to space limitations, we limit the majority of our discussion to the word and, to a lesser extent, the pseudoword results. We have demonstrated clear evidence of super-capacity processing of the word stimuli for nine of the ten participants. These participants are efficiently perceiving the whole word in comparison to individual letter perception. As mentioned earlier, evidence for the word superiority effect has been difficult to demonstrate with response times. These findings provide that evidence and thus agree with the majority of the word perception literature based on accuracy results. Based on comparisons across conditions, it is also clear that the

	Word	Pseudoword	Random	Upside-Down	Katakana
1	9.97^{***}	3.92^{***}	7.19^{***}	-2.62^{**}	-4.43^{***}
2	11.92^{***}	4.44^{***}	-0.73	-5.95^{***}	-10.02^{***}
3	8.19^{***}	-6.29^{***}	-6.88^{***}	-10.88^{***}	-12.34^{***}
4	0.13	-3.38^{***}	-7.34^{***}	-6.60^{***}	-10.58^{***}
5	0.79	10.70^{***}	-2.36^{*}	-6.27^{***}	-6.86^{***}
6	7.34^{***}	5.19^{***}	10.61^{***}	-2.58^{**}	-11.99^{***}
7	9.34^{***}	3.25^{**}	-2.27^{*}	-2.49^{*}	-5.78^{***}
8	7.17^{***}	7.84^{***}	4.68^{***}	2.86^{**}	-1.79
9	5.71^{***}	13.34^{***}	-8.43^{***}	-9.52^{***}	-7.37^{***}
10	3.88^{***}	2.45^{*}	-2.46^{*}	-7.44^{***}	-9.40^{***}
Group	20.38^{***}	13.11^{***}	-2.52^{*}	-16.28^{***}	-25.47^{***}

Table 2: Workload capacity statistical results for each participant. ($*** : p < .001$; $** : p < .01$; $* : p < .05$)

word perception was more efficient than non-word, upside-down word, and strings of Katakana perception, findings that again match with the results reported for accuracy (e.g., McClelland & Rumelhart, 1981). There is also evidence for a pseudoword superiority effect, another well established effect in the accuracy domain (McClelland & Johnston, 1977). Although the evidence was not as consistent as the word results, eight of the ten participants were super capacity for some time, with only two participants showing significantly limited capacity processing for most times.

There are multiple plausible explanations for the capacity coefficient results demonstrating particularly efficient processing of words. At least one of the assumptions of the UCIP model must have been violated, so we examine each of those assumptions in turn. Each of these violations have been considered previously for modeling the accuracy based superiority effects.

One assumption that may have been violated is that of independence. If there is any type of facilitation between the letter processes, each letter would be processed faster within a word which would explain the capacity coefficient values above one. There could be many explanations of this facilitation. For example, word processing mechanisms may in fact take advantage of the considerable amount of co-occurrence between letters in English. As is often observed, there are only a fraction of possible four letter combinations used for words and it would be surprising if we did not take some advantage of this reduction in uncertainty. This correlation between letters is an important part of how connectionist models explain the word superiority effect (McClelland & Rumelhart, 1981; Plaut, McClelland, Seidenberg, & Patterson, 1996; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001).

Many visual word processing models include a separate, phonological pathway. If a phoneme is activated as a possible interpretation of some letter combination, then it may in turn send positive feedback to those letters, speeding up their processing. Hence a phonological component of visual word processing could also lead to capacity coefficient values above one. Both the correlation between letters and the lack of a regular pronunciation of the non-words imply that these predictions are consistent with lack of evidence against the UCIP model of non-word processing. The phonological explanation is also supported by the evidence of a pseudoword superiority effect.

Another assumption of the UCIP model is that the letters are processed in parallel, with a separate detection of each letter. An alternative architecture that does predict capacity coefficient values above one is the coactive architecture. By pooling activation from each of the letters when processing a word, the word is processed much faster than if each letter is processed separately. A coactive architecture in this sense can be thought of as an extreme version of a facilitatory parallel model, in which all activation in each of the letters is shared. Many connectionist models of visual word perception assume a type of coactive architecture. In these models the activation accumulated in favor of a letter is immediately passed on to the word level. In this framework the type of parallel model assumed in the UCIP would not pass on any activation until the letter process is complete. A coactive architecture could also lead to violations of the assumption of unlimited capacity, so that seemingly more resources are available to each component when more components are present.

There were clearly individual differences present in these data, particularly in word and pseudoword processing capacity. This finding mirrors results reported in accuracy based studies (e.g., Reicher, 1969) and it will be an interesting extension of this work to compare the capacity measure to established measures of individual differences in reading.

Finally, we reiterate the importance of going beyond the simple ANOVA analysis of these data. Merely finding an ordering of the means in the string conditions says nothing about the relative processing efficiencies. For example, faster word processing than non-word processing could be due to the letters in "care" being relatively faster to process than the letters "rlkf". Workload capacity analysis, however, takes the processing of the components into account in estimating efficiency.

References

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind*, *11*, 63–65.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*, 204–256.

Houpt, J. W., & Townsend, J. T. (2012). Statistical measures for workload capacity analysis. *Journal of Mathematical Psychology*, *56*, 341–355.

LaBerge, D., & Samuels, S. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*, 293–323.

McClelland, J. L., & Johnston, J. C. (1977). The role of familiar units in perception of words and nonwords. *Perception and Psychophysics*, *22*, 249–261.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter percpetion: Part 1. an account of basic findings. *Psychological Review*, *88*, 375–407.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*, 56–115.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *Quarterly Journal of Experimental Psychology*, *55A*, 1339–1362.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, *81*, 274–280.

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial and coactive theories. *Journal of Mathematical Psychology*, *39*, 321–360.

Townsend, J. T., & Wenger, M. J. (2004). A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review*, *111*, 1003–1035.

REAL-TIME GATING OF CYCLIC BRAIN ACTIVITY: EVIDENCE FROM MODULATIONS OF BINOCULR RIVALRY INDUCED BY RAPID STIMULUS INTERRUPTIONS†

H.-G. Geissler[1]*, V. Vanagas[2], A. Svegzda[2], R. Bliumas[2], R. Stanikunas[2], H. Vaitkevicius[2]**
[1]University of Leipzig; [2]University of Vilnius
* hans-g.geissler@rz.uni-leipzig.de ** Henrikas.Vaitkevicius@ff.vu.lt

Abstract

According to widespread opinion, perceptual-cognitive information processing is founded on systems of mutually tuned neural oscillations flexibly adapting to environmental constraints and task demands. From directly observable phenomena ubiquitous in the brain, this view seems to be well-supported. However, it ignores important prerequisites of stability, selectivity and protection against disturbances. Regular timing patterns surfacing in resonant states of the brain suggest that they are accomplished by delay-based gating mechanisms whose action in support of precise temporal coordination hardly ever comes directly into sight. Exploring the effects of rapid stimulus interruptions (25 to 125 Hz) on binocular rivalry, we demonstrate – for the first time for such a large range – fine-grained cyclic modulations that are evenly spaced in time rather than frequency. Our results support the view that recurrent chains of neural delays play a leading role in the selective amplification of preferred cycles. This agrees with range constraints specified by the time-quantum approach, TQM.

The claim that the bodily basis of perception and consciousness resides in cyclic carrier processes within the nervous system dates back to the very beginnings of psychophysics (Fechner, 1860). What was a speculative idea in those days has led after the discovery of periodic fluctuations in potentials recordable from the scalp (Berger, 1929) to a rich literature on correlative relationships. It is only during the last few decades that various testable hypotheses on the causal roles of synchronous oscillatory activity in information processing have become the focus of extensive physiological research. Among others, the idea has received prominence that binding of segregated pieces of sensory information into unique percepts is achieved through synchronous oscillations of populations of neurons (Eckhorn et al., 1988; Gray et al., 1989); cycles of fast oscillations in the Gamma and Beta ranges were proposed to result from segmentation of longer cycles in the Alpha or Theta bands (Lisman & Idiart, 1995; cf. Sompolinsky & Tsodyks, 1994, for an analogue on the neuron level); more recently, a mechanism for recoding amplitude information into phase positions within the gamma cycle was suggested (Fries et al., 2007, for a brief review). Also, great efforts have been undertaken to disentangle the complex functional interrelationships of activity in different frequency bands in early stages of cognition. A salient example is the inhibition-timing hypothesis (Klimesch et al., 2007; Klimesch, 2011) claiming that ongoing alpha activity and early evoked visual (P1) potentials form an integrated functional system selectively channeling initial memory access.

As part of those developments, there arose and took root two closely related, but more unspecific suggestions that because of their universal character cannot be tackled on the same, primarily inductive basis. One is the supposition that frequency and phase of neural oscillations, or compounds of them, may generally play the roles of fundamental parameters for transient coding and processing of information throughout the brain as a whole. The other assumption is more of a background idea, namely the widely shared, though unproven, belief that all forms of functionally engaged oscillations of cooperative and competitive neuronal mass activity can ultimately be captured as facets of the behavior of non-linear oscillators.

Hopes that these fairly undifferentiated claims could before long be made exact within readily available system-theoretic frameworks and thereby converted into testable predictions have so far not come true. In this paper, we adopt a deductive alternative capitalizing upon psychophysical evidence. As we will show, law-like invariance properties inferred from fine-grained perceptual-cognitive response patterns bear directly on the phase-frequency issue. Yet, while suggesting a beautiful simple compatibility of phase-frequency coding in absolute terms, the same principles simultaneously challenge plain oscillator intuitions. Specifically, the observed regularities point to a less easily detectable second type of cyclically operating real-time mechanisms that tend to suppress detrimental oscillatory driving and generate states of fine-tuned transient quantization vital for stability and sustained multi-leveled process organization. In the second part of this paper we will for the first time report evidence in support of this view from a critical test in an experiment in binocular rivalry, the alternation of perceived images occurring when the eyes are presented with incompatible visual patterns (see Blake, 2005, for an introduction).

Because of integer-ratio relationships between adjacent components, the spectrum-like discrete patterns from which the key invariance properties can be extracted will here be referred to as "quantal structures". As a compact basis of accumulated evidence, we relate to the so-called Time Quantum-Model, TQM − not a model in the customary sense, but a condensed, continuously updated interpretation of laws and numerical invariants inferred from a larger collection of quantal structures (see Geissler, 1987, 1992, for early versions).

As a prerequisite for treatment of the issues at hand, a brief recapitulation of essentials of TQM appears imperative.

Basic hypotheses and TQM core structure

To start with the general rationale of the enterprise: Unlike common inductive strategies that translate empirically established relationships into directly testable hypotheses or into models from which directly testable hypotheses are derived, the TQM approach translates the invariant properties it sets out from into currently not directly testable general hypotheses that are treated as axioms and are not required to form a complete system or model. TQM itself consists in a revisable integrated interpretation of these hypotheses in terms of underlying cyclic processes. On a trial basis, this core structure is implemented for the phenomena under investigation in order to derive testable predictions. Thus the essence of TQM lies exclusively in the predictive power of the implementations of the hypotheses adopted, which allow deductions that are subject to empirical corroboration or falsification.

Among the three most basic hypotheses entering into TQM there are two that root in regularities that quantal time structures not only share with *non-quantized* time dependencies, but also with psychophysical ratio-scaling data that are obtained for *non-temporal* continua of the character of intensities (Teghtsoonian, 1971). Most fundamental is the observation that quantal time structures in terms of the proportions between the largest and smallest values across limited ranges do not exceed upper bounds of 20 to maximally 30. In the renowned "multiplication table" of Latour (1967), for example, the lowest value is located somewhat above 9 ms and the highest (concluded) one at 220 ms, which yields a ratio close to 24. For TQM, estimates of the obtainable maximum have led to Range Hypothesis (**H1**), maintaining that *internal representations of time form ranges such that the ratio of the largest to the smallest possible value within a given range is a constant* $M \simeq 30$. Note that this figure agrees fairly closely with the constant established by Teghtsoonian for judged magnitudes (Teghtsoonian, 1971; see also Teghtsoonian & Teghtsoonian, 1997).

The second basic regularity relates to an empirical law ubiquitous in psychophysics: Weber's Law, which was generalized for time as "Scalar Timing" (Gibbon, 1977). The corresponding TQM hypothesis for internal representations was first stated in Geissler (1985). Its quantitative specification, first derived by Teghtsoonian (1971) in terms of judged magnitudes, amounts to the Dispersion-Progression Hypothesis (**H2**) saying that *along a full quantal range the Weber fraction has a constant value* $C \cong 1/30$ *(Ekman's Constant").*

H1 and H2, if stated in terms of judged magnitudes, do not straightly transfer to physical values with possible equivalents in physiological observations. But their time-related versions do. Yet, due to the relational nature of H1 and H2, absolute measures of time are still lacking. The key hypothesis through which TQM fills this gap derives from a third inherent regularity of quantal structures, namely spacing of components approximately on a uniform lattice of possible values. This expresses itself in Greatest Common Denominators that according to H1 increase in linear progression with range size, starting from a smallest absolute value corresponding to the shortest range observed. A landmark on the way to the disclosure of this intriguing regularity is represented in work on visual pattern recognition by Vanagas, Balkelite, Bartusjavicus, & Kirvelis (1976; cf. also Vanagas, 1994, 2001). In this study, due to an unusually minute gradation of stimulus presentation times, small processing steps of, on average, about 9 milliseconds could for the first time be made visible in the percentage-correct functions. Critical intervals of that duration or small integer multiples of it have later repeatedly been also found in RT-based analyses (e.g. Puffe, 1990; Bredenkamp, 1993; Petzold & Edeler, 1995). The seemingly universal role of an even shorter unit of half that size, i.e. of ~4.5 ms (in the following, Q_0) became apparent from a reanalysis of larger samples of quantal structures from different paradigms by Geissler (1987). Beyond its demonstration as Largest Common Denominator, most convincing evidence of its real existence so far comes from analyses of small fluctuations under near-identical conditions (cf. Geissler et al., 1999; Kompass, 2004). As a quantitative postulate about internal representations, the modular property described was stated as Time-Quantum Hypothesis (**H3**): *Possible periods are integer multiples N of modular units* Q_q (H3a), *which in turn are integer multiples q of an elementary smallest unit* Q_0 (H3b).

For illustration of how the hypotheses relate to TQM, consider Figure 1. In the wave scheme at the top, the modular quantum unit according to (H3a) is implemented as a "fuzzy"bipolar sinusoidal cycle resulting from superposition of carriers of slightly different cycle durations (cf. Geissler, 1985, 1987). Note that, after initial synchronization (rectangular

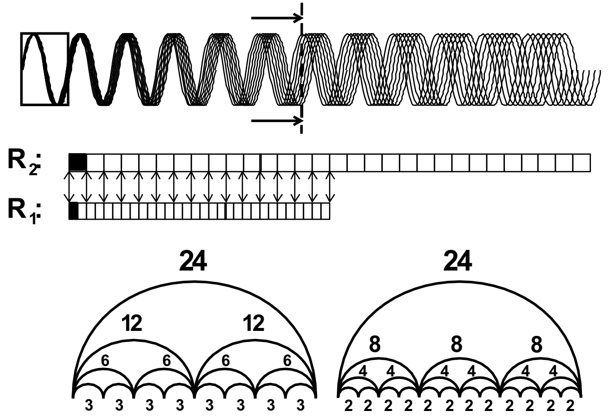


Fig.1. Illustration of the TQM assumptions H1, H2 and H3. Note that in the picture at the top random phase differences are depicted as eight waves of equal distances.

frame at the top left), in accordance with H3 phase jittering spreads linearly as time elapses after initial synchronization (dashed vertical line with arrows). It does so up until the "coherence limit" of N_{max} cycles (here arbitrarily assumed to be 12) is reached where it completely erases detectable periodicities and thus no phase coding of information would anymore be possible. Consequently, N_{max} is the only resulting limit that can be identified with the upper range bound M according to H1. Note that this identification brings something new beyond H1 and H2 as such. It follows automatically that M and C can be reduced to only one constant, the uncertainty interval of the operative modular unit. When assuming uniform phase distribution, this is tantamount to the testable prediction that $C \sim 1/M$ should hold. The approximate inverse relation between the estimates for M and C, never taken notice of in the literature, fits in well with this prediction.

It is now easy to capture an important consequence of H3b: Unlike the other hypotheses that are stated in relative terms, Q_0 provides a basis to specify ranges in absolute terms, which is big news in psychology. A straightforward consequence is the existence of an absolutely shortest range R_1 of possible multiples of Q_0. When adopting the ideal estimates $Q_0 = 4.57$ ms (for adults) and M = 30, this "base range" extends from durations of 4.57 ms to 137.1 ms. A compelling requirement for mental timing as a whole is that ranges of different absolute extensions up to the respective Weber resolutions must be compatible within sub-ranges of overlap. In the middle part of Figure 1, this property is illustrated for R_1 and R_2 by bidirectional arrows connecting the lower chain to the next larger compatible chain above it. Accordingly, integer-ratio relations should hold between ranges of different sizes in order to meet the compatibility condition. By generalization, there follows a proper cascade R_1, R_2, R_3,.,., etc.,. of admissible ranges resulting from R_1 by magnifications through integer factors.

Motivation of the experimental inquiry

That much about TQM; let us now return to our main objective: What have temporal range structures to do with phase-frequency relations and their connection to oscillatory mechanisms, the two central issues of this paper? For phase and frequency, the answer is straightforward: According to H1, quantized cycle times are evenly spaced by physical measures of time. If, for the sake of simplicity fuzziness is ignored, this property implies a simple relation between cycles and phase positions. Possible phase positions can, across all admissible cycles, be expressed as multiples of the same modular unit as the cycle itself within which phase positions are marked out. As one can easily imagine, this relationship enormously simplifies extraction and computation of complex information across different frequencies. By contrast, there would be no such simple rules on which processing can rely if cycles could take on any value or were uniformly spaced in terms of frequency.

What about oscillatory mechanisms? Superficially, it might appear that range structures of cycles uniformly spaced in time logically contradict customary oscillator concepts, as they typically ascribe an analogous privileged role to multiples of natural frequencies of oscillators, which amounts to segmentation of largest rather than concatenation of smallest cycles. However, reality is much subtler. In fact, analyses of inner regularities of quantal structures such as were first described by von Békésy (1936) for low-frequency sound and by Latour (1967) for eye movements and their relations to EEG frequencies have revealed exactly that *hierarchical* segmentation property. The hypothesis that it holds simultaneously with the time-quantum assumption H3 is among the earliest suggestions of TQM (Geissler, 1985). For full ranges, i.e. for modular multiples $1 \leq N \leq M = 30$ as hypothesized, it can be stated as follows:

<u>The Double-constraint hypothesis</u> (H4)

Admissible processing cycles are both integer multiples of the respective operative modular cycles and integer fractions of larger multiples if there are any.

Thus H4 agrees in one basic respect, the “segmentation property”, with common oscillator concepts. However, it differs from them by two properties with crucial consequences for the main issues of the present article:

(A) As illustrated at the bottom of Figure 1 for two different segmentations of a macro cycle of 24times a modular unit, the potential manifold of possible segmentations becomes narrowed down to a strictly limited set of hierarchies. More exactly, ignoring fuzziness, hierarchical segmentation can be reduced to decompositions following the *mathematics of alternative ordered factorizations of integers* 1 to M.

As a consequence of (A), the above indicated uniform relations of within-cycle phase positions across admissible cycles translate into concrete relations between cooperating, hierarchically phase-coupled carriers of cyclic activity of different frequencies.

(B) Different from common biological concepts, there are no fixed oscillators or clock mechanisms implied. Instead, each multiple $1 < N \leq M$ can in principle assume the position of the top element in one or more segmentation hierarchies.

Note that (B) thus delimits the manifold of available options and does not say anything about their employment in the representation and processing of information. Accordingly, TQM accounts for task- and stimulus-specific processing through fast progressive selection from a large number of initially co-activated options, functions that in common oscillator implementations are ascribed to adaptive tuning to stimulus conditions and task demands.

So far the strongest empirical evidence in support of (A) and (B) that offers a direct link to the objectives of this paper can be seen in the success of two predictions. One relates to a long-known global physiological regularity, the other to stochastic properties of a class of ambiguity phenomena. Due to lack of space, we can here provide only a sketchy description: The first prediction springs directly from property (A): As a consequence of prime factorization, the set of possible multiples $N \leq M = 30$ falls apart into four subsets of mutually disjunctive elements. In application to R_1, this decomposition reveals a perplexing congruence with the empirical definitions of EEG bands in terms of frequency values (see Appendix). What is important in the present context is that the even spacing of EEG bands on a logarithmic scale corroborates the even-spacing assumption for possible cycle durations.

The second prediction relates to superficially irregular multimodal distributions of breakdowns of apparent motion (cf. Geissler, 2009). The relevant predictions derive under the premise that for the weakly-constraining conditions of the experiment all possible options are activated with equal probability. Although modified variants seem possible, the high correlations found for more than 20 degrees of freedom should hardly allow for completely different explanations. What bears on the objectives of this paper is that the predictions rely on the above progressive constraining principle and thus contradict plain oscillator notions.

While the evidence quoted in favor of the quantal-timing claims is fairly encouraging, one may still doubt their generality. Against the power of the EEG prediction one may object that it is not a forecast of a hitherto unknown phenomenon but a re-production of a familiar structure that as such is suspicious of depending, at least partly, on technical conventions. An obvious limitation of the second example is its limited scope: Due to inherent constraints of the phenomena investigated, Beta and Gamma Motion, predictions apply to ISIs up to 250 ms, a bit below the upper bound of R_2, which corresponds to the lower limit of the Theta band.

Thus it remains desirable to demonstrate broad validity of the assumptions more directly, for a much larger range of variation of temporal parameters and within a paradigm building on robust effects providing a strong test. Another highly desirable requirement is that the paradigm employed should psychophysically and physiologically be well studied – quite in accordance with Fechner’s ideal goal of establishing “Inner Psychophysics” through linking psychophysical and physiological evidence.

Proposed mechanisms and choice of paradigm

A goal thus set raises the question what the physiological basis of the curious double-constraint condition might be. A cue to answer the question is provided by a comparison of different interpretations of the same psychophysical principle as illustrated at the bottom of Figure 1 (cf. Geissler, 1994). When considering that scheme, the irresistible primary interpretation is one in terms of two alternative hierarchies marking cycles of four phase-coupled oscillations whose largest members are of equal duration. However, alternatively, the hierarchies can be conceived as two different segmentations of one and the same chain of elementary cycles or intermittencies, x, into chunks according to

((((xxx)(xxx))((xxx)(xxx)))((((xxx)(xxx))((xxx)(xxx)))
or ((((xx)(xx))((xx)(xx)))(((xx)(xx))((xx)(xx)))(((xx)(xx))((xx)(xx)))).

The important difference between the two modes is that implementation in terms of oscillations in either case involves participation of four different oscillations, whereas in the chunking interpretation the resulting structure somehow "consists" of sub-chunks as constitutive parts: if one of it is dropped the total chunk is no more of the same length. In fact, while prediction (B) implies transient generation of full hierarchies, it is inconceivable how (in the example) four different oscillations could be generated fast enough and brought to precise phase coincidence. Moreover, in the face of well-known properties of non-linear oscillators, it remains a puzzle how the configuration could be kept stable and frequency floating avoided. Quite differently, whatever chunking exactly means, in the second version stability is not a question of precise coordination, for sub-chains are constitutive parts of the whole. Also, it appears no principal problem to add new sub-chains or replace a particular decomposition by another one.

Considerations like the latter have led to the suggestion that the neural basis of temporal quantization lies in precisely timed chains of neural delays (cf. Kompass & Geissler, 2003; Kompass, 2004) or "synfire chains" as first postulated on the basis of findings of super-precise timing on the neuron level in the forebrains of monkeys (Abeles et al., 1993). The resulting hypothesis thus is that the physical basis of the quantal property is given by recurrent neural delays that by imposing their constraints upon neuronal mass activity co-determined by informational constraints control the emergence of transient resonant states of cyclic activity. For details of possible implementations the reader is referred to Kompass (2004).

The mechanism assumed gives rise to expectations on which experimental tests can be based. The functioning of the "invisible" delay mechanism should be robust even under quite artificial conditions. This should make possible psychophysical tests over much larger ranges of variation of temporal parameters than can be realized within the confines of spontaneous temporal organization induced solely by task-related performance. Other experimental situations to be taken into consideration are given by a large class of driving procedures in which normal performance is altered through enforcing external rhythms either by cyclic presentation of stimuli in addition to task-related target stimuli or by cyclic variation of the intensity of the target stimuli. Obvious physiological parallels for possible comparisons are stimulus regimes as are employed in studying steady-state EEG responses. In psychophysical timing research the application of driving techniques has received a major boost through pioneering work by Treisman and coworkers demonstrating pronounced driving exerted by series of clicks upon judged durations (e.g. Treisman, Faulkner, Naish, & Brogan, 1990). An important further step was made by Burle & Bonnet (1999, 2000). Crucially, by fine-graded variation within in a small interval of frequencies, these authors found in Sternberg's task and in a Simon paradigm across several conditions the same modulation effect in form of a near-sinusoidal wave. Different from their expectations, this function cannot, however, be explained by physical driving acting on an oscillator controlling timing, because it again steeply crosses the axis where it should relax to it. Important for interpretation within the framework of TQM was the finding of Kompass (2004) that the distance of the outer knots of the wave agrees quite closely with Q_0, later complemented by the observation of their agreement with multiples 10 and 11 (Geissler, 2009).

In order to explore the even-spacing assumption in a large range of presumed intermittencies, binocular rivalry was adopted as paradigm. To demonstrate fine-grained modulations of Dominance Time (DT), here defined as the period of time during which one of the mutually excluding images is continuously perceived, stimuli were cyclically interrupted. This choice of conditions is a challenge from the viewpoint of the Q_0-related modulations described, because DTs are of the order of seconds as compared to latencies never exceeding a few hundred milliseconds obtained in the RT-paradigms so far studied. Inclusion of high interruption rates is also a challenge, because for rates above the so-called critical fusion frequency (CFF) non-monotonic effects have never been observed. To avoid complexities concerning differential effects of light intensity, the lower bound of interruption rates was set at 50 Hz. To permit a clear decision about the hypothesis of even spacing of fine-grained modulations in time, the range of driving rates was extended up to 125 Hz.

Working hypotheses

Although as yet very little is known about the precise nature of the observed modulations, TQM allows for rough expectations. In the experiment by Burle and Bonnet (2000), RT effects of task-related stimulus variation of maximally 150 ms, i.e. of about the extension of R_1 corresponded to driving modulations of ~5 ms. Kompass' observation of a driving cycle roughly equal to Q_0 perfectly fits into this frame. Thus these figures suggest that driving causes essentially stochastic delays or accelerations of no more than one quantal epoch within R_1. For binocular rivalry, the respective critical intervals are of the order of one to several seconds and they comprise three generally accepted stages all of which may be sensitive to cyclic driving. As a consequence, we may expect driving effects ranging from minimally ± 50 ms up to ± 100 ms and more. In addition, we should expect a mixture of effects ranging from contributions of Q_0 and $2Q_0$ to possibly $16Q_0$. Still, if the gating assumption is valid, near uniformity in time should hold in every case.

Experiment

Subjects

Four subjects of normal or corrected-to-normal vision participated in two to three sessions, in the following indicated by DN, RB, NK, and LO. For simplicity, cardinal numbers of sessions and eye sides (L, R) will be attached to these acronyms. Thus, for example, DN3*l* relates to data obtained for subject DN, for his left eye in the third session.

Apparatus, stimuli and design

In a completely darkened room subjects were exposed to transparent slides mounted in a Wheatstone stereoscope and lighted by PC-controlled white LEDs (of ~60 cd/m^2 luminance). Black bars (1.7° × 0.6°) were presented to the left and right eye against white backgrounds at 45° counter clockwise and clockwise orientations, respectively. The slides were lighted asynchronically with equal presentation and interruption times with an accuracy of ~5 μs. Subjects were asked to press a button as soon as an alteration of the perceived bar orientation occurred. Responses were automatically registered and Dominance Times computed for each eye separately. One session consisted of 17 blocks, each for one fixed value t_i randomly chosen from presentation times spaced at millisecond distances from $t_1 = 4$ to $t_{17} = 20$ ms. A block lasted about three minutes. For two participants (DN, NK), the number of sequential key pressings, in the following referred to as trials, was fixed to 50, for each eye side. For the other two participants, as usual in similar experiments, a time limit was set leading to variable numbers of trials between 37 and 50. Blocks were separated by short breaks of three to five minutes of rest.

Results

Fine-grained systematic modulations could be confirmed for all participants and sessions with block means of DT_i differing significantly at least between opposite extremes. Most strikingly, absolute elongations were found to be up to 50 times as large as those established in other driving paradigms. However, at the same time large differences resisting any type of standard analysis were evident not only in apparent periodicity but also in absolute elongations, between subjects and eyes as well as within subjects between sessions. In a process of careful stepwise exploration, uniform dynamic properties could be dissociated from striking unpredictably occurring dynamic-shift phenomena of yet unknown origin. In that way basic driving characteristics could be isolated. A key to reach this goal was provided by cases in which clear periodicities of the anticipated type were obvious. A pronounced example is shown in Figure 2: While the data from the first session of subject DN (DN1) in the left panel show a base level about three times as high as those from the first session of RB in the right panel, a strong periodic modulation is found for the left-eye side of RB. With elongations of maximally 0.4 s to both sides, it covers about one third of the entire range. Two clearly visible bipolar waves point to a cycle as a function of presentation time of 4 to 5 ms, which is of the expected order.

Another finding contributing to the disclosure of basic dynamic effect properties is shown in Figure 3. Namely, when computing the averages across sessions and eyes for the subjects whose first-session data are shown in Figure 2 it turns out that in either case about the same periodicity appears as the one seen for the left-eye data at the right of Figure 2. This holds with one striking exception: the completely reversed trend within range t_1 = 4 ms to t_6 = 9 ms. For convenience of illustration, in Figure 3 this section is mirrored at the overall mean which is put to zero and contrasted with the rest of the function using open diamonds and dashed lines. Comparison with a sinusoidal function of 4.5 ms cycle duration (dotted lines) makes two things obvious: First, both the empirical functions are of roughly that period and, second, the overall course in both cases is that of a damped oscillation becoming increasingly flatter for presentation times t_i > 14 ms. Via generalization encouraged by similar observations, these findings lead to two modifications of the original hypothesis on the course of the induced effects, which was derived from the trend obtained by Burle and Bonnet for a much smaller interval of driving rates: (1) Amplitudes may vary and even change signs depending on organismic state variations. (2) They tend to decrease and get more variable as stimulus cycles increase. Note that knot positions along the t axis remain invariant.

Results like those raise the question whether a periodicity of this order is inherent to the entire data set. Motivated by experience from work in color-space representation (Fomin, Sokolov & Vaitkevicius, 1979), factor-analytic techniques were applied.

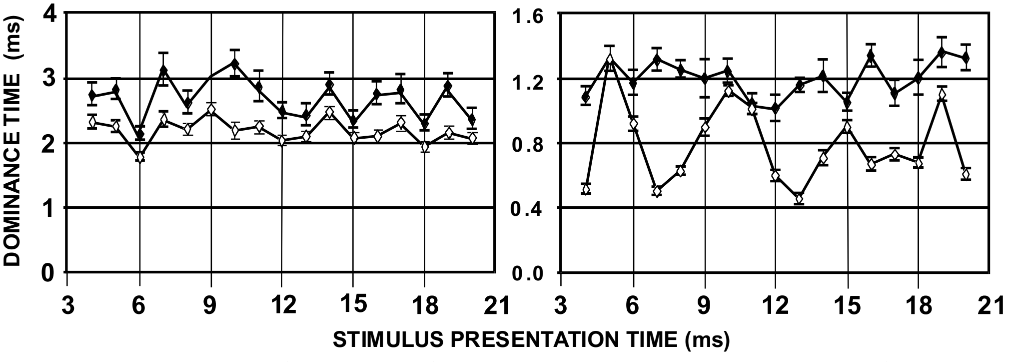


Fig. 2. DT for DN1 and RB1 plotted against flash duration. Vertical bars denote 95% confidence intervals. Filled and open symbols denote left eye and right eye data, respectively.

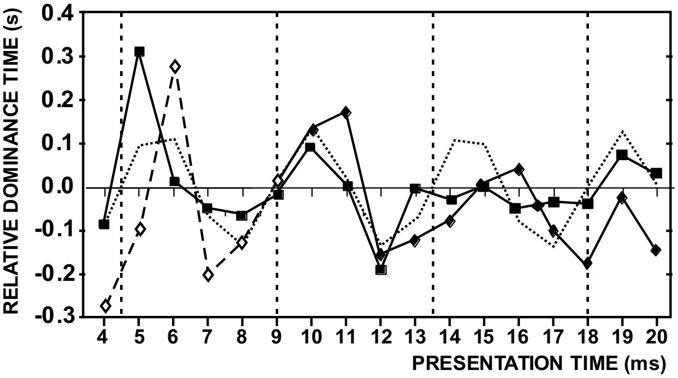


Fig. 3. Mean DT trends across sessions relative to the overall means for subject DN (open and filled diamonds, see text) as compared with subject RB (filled squares).

To eliminate the dominant influence of the large base-level differences between individuals, session means of DT were subtracted from block means for each eye separately. From the resulting profiles of relative DT values a 20 × 20 correlation matrix was computed and subjected to Principal Component Analysis (PCA). From the four main factors accounting for 80% of the total variance, the third in the order of variance contribution (F3) turned out to exhibit a strikingly periodic loading structure (Table 1). Rotation of the factor space revealed no crucial change in this patterning. Also, the corresponding coordinates on a dimension obtained through a complementary DMS analysis replicated essentially the same structure. Fitting a sinusoidal function to the F3 loading profile enforcing zero phase shift and using the criterion of maximum correlation revealed a cycle of ~4.45 ms.

t_i (ms)	4	5	6	7	8	9	10	11	12
Load	-0.367	0.703	0.609	-0.638	-0.504	-0.220	0.120	0.011	-0.385
t_i (ms)	13	14	15	16	17	18	19	20	
Load	-0.155	0.085	0.400	-0.118	-0.148	0.158	0.414	0.160	

Table 1. Load of factor F3 as a function of presentation times t_i (in ms)

Although the regularity of the F3-load structure is robust enough to appear as a cyclic modulation even in the one-factor solution obtained when entering analysis with unreduced data, the result is not fully satisfactory on two major counts. First, F3 absorbs only about 20 percent of the variance covered by the remaining factors. Closer inspection in accordance with (1) and (2) above reveals as likely reasons volatile fluctuations and a steep decline of amplitudes with increasing stimulus presentation times. Both features will drastically attenuate correlations and as a result the contribution to overall variance. The second reason consists in what appears an irregular structure of the eigenvectors of the remaining three factors. The expected contributions of Q_0 and its multiples larger than $2Q_0$ are not visible.

To find a solution, we have to first assign the above estimate of cycle duration in terms of presentation time to a hypothetical intermittency in the brain. The standard answer suggesting itself is based on the observation first reported by Sherrington (1904) that effects of alternating and synchronous monocular interruptions do not differ for high presentation rates. The common conclusion is that only the monocular rate is relevant. In application to factor F3, this implies that the underlying physiological cycle should be of the duration ~2 × 4.45 ms = 8.9 ms, i.e. of the order of the "Vanagas quantum" mentioned above. An intuitive further consequence would be that that value marks the lower limit of resolution, which would be at variance with former findings such as those of Burle and Bonnet (1999, 2000) amounting to an operative lower limit of ~4.5 ms, i.e. of the order of Q_0.

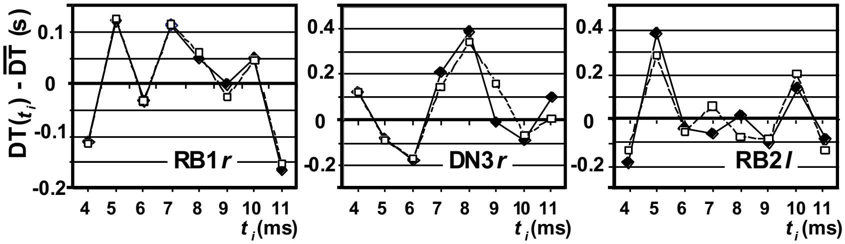


Fig. 4. Fits (open squares) of (1) to irregular data (black diamonds) for RB1r, DN3r, and RB2l. Correlations are, from left to right, 0.99, 0.81, and 0.80.

To resolve this puzzle, recall that the equivalence assumption of alternating and synchronous stimulation is based on monotonous dependencies and thus does not include periodic modulations even of ~9 ms. In this perspective, therefore, stimulus on- and off-sets across both eyes can be very well established at central parts of the brain rhythms of twice that rate. Note that for intermittencies thus short effect variation between sessions as described by (1) may even be more volatile and decline with increasing presentation time correspondingly steeper. This would explain the lack of any direct trace in the factor structures and call for regression procedures as a viable alternative. Yet, instead of standard procedures, complex iterative strategies will be required; because fixed zero crossings at integer multiples q of Q_0 seem to be the only reliable constraints for sinusoidal approximations. As a first move in this direction, session means for the first eight presentation times from $t = 4$ to $t = 11$ms, corresponding to the segment of highest correlations with F3, were fitted to linear combinations of the form

$$\psi = \Sigma_n a_n * \sin(2\pi * t/2^n * Q_0) + b_0. \qquad (1)$$

In addition, the empirically founded doubling rule $q = 2^n$ and an upper limit of n = 3 were adopted as pragmatic constraints. Furthermore, the estimate $Q_0 \cong 4.45$ ms obtained from fitting the F3 eigenvector was used for all computations – a rather strong simplification. Note also that b_0 and the expressions for $n = 2$ and 3 may in effect merely play the role of functions taken as customary for non-linear trend reduction. Due to possible differential variations within sub-cycles, no particularly good quantitative fit was to be expected. However, substantial correlations of, on average, r = 0.62 justified the attempt. Since in ten of the twenty cases the contributions of Q_0 turn out to be larger than those of $2Q_0$, structural agreement with the data can rarely be attributed to the inclusion of higher-order multiples. Above all, the excellent fit of particularly irregular structures of which three striking examples are shown in Figure 4 speaks for a genuine role of Q_0.

Discussion

Basic findings

Figure 5 illustrates the probably most important result of the foregoing analysis. In the panel at the top, the plot against *presentation time* of the F3 load structure (black diamonds) together with the computed sinusoidal approximation (open squares, dotted lines) makes the basically even cycle-spacing apparent. These modulations established for a range of rate variation as large as 25 Hz to 125 Hz clearly speak in favor of the real-time nature of the presumed quantal mechanism, despite some imperfections of the detected regularity. By comparison, the panel below, in which the same data are re-plotted as functions of the reciprocal *presentation rate,* is striking because of an accelerated compression toward the right side. It thus reemphasizes on a purely qualitative basis the relevance of the even-spacing feature getting visible in the time domain.

In conclusion, this basic finding supports the notion that two different principles are at work in joint action, shaping the temporal regimes of information processing throughout the brain. Of these, the one reflected by the driving structure seems to ensure under natural conditions precise and undisturbed timing by gating and supporting the stable structuring of wave-like oscillatory activity that is governed by the second principle, the segmentation of longer periods into smaller ones.

We appreciate a similar earlier idea of dual pacing suggested in an insightful paper of Burle and Bonnet (2003). However, an adequate comprehensive implementation must go beyond clock-like mechanisms as the basis of cyclic timing. Neither the progression of maxima, as would have to be expected in accordance with Treisman et al. (1990), nor that of the zero crossings in the middle of each full cycle as considered by Burle and Bonnet agree with harmonics of a wherever located natural oscillator frequency. In addition, modulation amplitudes are the higher the higher the driving rates, a horror to any plausible pure oscillator implementation. But this feature fits nicely in with TQM if one assumes that a driving rhythm elicits fuzzy carriers of cycles corresponding to the boundaries $N \times Q_0$ and $(N+1) \times Q_0$ of the interval into which the driving cycle falls, because the amplitudes of the superimposed carriers added together will be the higher the smaller the zones of fuzziness.

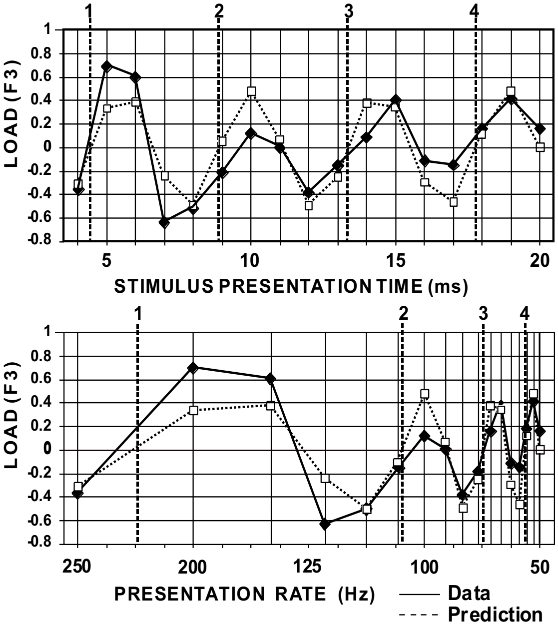


Fig. 5. Top: F3 loading structure (filled diamonds) and predictions from fit of a sinusoidal function (open squares), plotted against presentation time. Bottom: same, against rate.

We can only in passing mention here that the evidence supporting the validity of H1 and H2 not just for *time* but also for *intensity* suggests that the assumed time-based principle also applies to intensities in the brain. This agrees with the mechanism described in Fries at al. (2007) for recoding spike-train coding of intensities into temporal phase coding within the cycle of gamma activity. One may thus speculate that spike trains are intimately related to the second principle of oscillator-like action.

In accordance with the focus of this paper, we have so far concentrated on relational properties. In a more general perspective, it is no less important that our data support basic tenets of TQM with respect to the order of magnitude of the elementary time unit assumed. Whether one agrees to the proposed precise value of the lower limit and its assumed role or not, the cycle identified in the F3 load structure and the results of the above regression analysis argue against the theoretical relevance of intermittencies as large as some tenths of milliseconds, which are often adopted as fundamental units. Instead, they argue in favor of smaller units far below conscious discrimination. From the TQM perspective, it is this subtle graining in the first place that provides room for fast processes whose automatic underpinning renders possible conscious cognition.

The significance of driving procedures

Mere confirmation of the even-spacing hypothesis does by far not exhaust the returns from the experiment. Of significance for a deeper understanding of the assumed quantal-timing mechanism are two general dynamic properties of driving effects already found in Burle and Bonnet's data and again confirmed in the present experiment: First, it is the interpolated zero-crossings that coincide with range-specific multiples as predicted, whereas the space between them is filled with up-and-down variations. Qualitatively identical bipolar courses found for driving rates from 25 to 125 Hz indicate robust and invariant functioning of one and the same mechanism along the whole range and probably beyond. This requirement is plausibly met by the delay-based implementation hypothesized in Geissler & Kompass (2003) and its possible neural specifications (cf. Kompass, 2004).

The second feature of importance in the present context springs from comparison with features of spontaneously emerging quantal structures. What is characteristic of the latter is selective prominence of subsets of quantal multiples in accordance with H4. In fact, in driving profiles such preferences are completely lacking. In other words, driving rhythms seem to completely override this constraint and instead enforce themselves as the dominant rhythms. As we have seen, the working hypothesis that this is due to superposition of elicited fuzzy carriers seems to provide a good description of this fact. It remains an open question how this can be expressed in terms of the assumed timing principles.

Are there benefits for research in binocular rivalry?

To complete the picture, we should address the question of possible benefits for research in the field of binocular rivalry (BR). Two facts through which our results bear on current theorizing on the phenomenon are fairly obvious: First, none of the accounts of BR have so far led to the expectation of fine-grained systematic driving effects as those reported here. Second, our findings give testimony to temporal characteristics in the millisecond range by which different individuals agree within margins of a few percent. To the best of our knowledge, nobody working experimentally in the field of BR has as yet even tried to demonstrate such commonalities. Too large are the individual differences in BR to encourage such an endeavor. Yet, as is well known, failing recognition of a regularity inherent to a complex structure may render detection of others difficult or impossible. In this context, it is of interest that TQM not only offers estimates of the invariants to be accounted for, but also gives advice on how individual differences can be included without loosing the foothold of invariance. Just to mention two cases: In the wave scheme of Figure 1 ideal initial synchronization can be replaced by phase spreads specific to individuals, causing a reduction of the observable coherence limit to an "operative" value $M^* < M$. Similarly, individuals may differ in the Q_0 modular multiples adopted for otherwise identical processing stages which could explain the observed large base-level differences. There is no space here to explore those or any further options. Suffice it to say, even combinations of such modifications remain testable provided the basic invariants are left untouched.

There are indeed various indications giving evidence that, beyond the observed fine-grained modulations, the known characteristics of BR are in close accord with fundamental invariants as are suggested by TQM. Most straightforward is the apparent correspondence between shortest ranges according to TQM and persistence times of early stages of information representation and processing: Referring to analogous earlier findings by Engel (1970) O'Shea and Crassini (1987) report a ubiquitous upper bound of about 100 ms common to stereopsis and binocular rivalry and a persistence of a second stage of integrated binocular information of ~300 ms. According to TQM, information can be first represented in R_1 with a maximum precision of ~4.5 ms. When the coherence limit is reached at ~135 ms, this precision is lost and cannot be recovered. The next finest possible resolution is ~9 ms and the pertaining ideal upper limit is ~ 30 × 9 = 270 ms. These figures agree quite closely with the empirical persistence characteristics and therefore suggest congruence between assumed range structure and physiological stage organization as a reasonable hypothesis. To further check this hypothesis a comparison with estimates obtained by modeling full DT distributions would be desirable. As a first attempt, we here refer to a model of Manousakis (2009) because of its unprecedented data fit with a small number of parameters, which does not imply we share his quantum-physical interpretation. In its mathematical form, the model closely resembles an earlier suggestion by Atmanspacher (e.g. Atmanspacher et al., 2004). Crucial for testable predictions is the theoretical forecast of a slowing-down of state transitions, the quantum "Zeno Effect". Interestingly, fit of three well-documented cases yields intervals of 300, 280 and 100 ms that can be conceived of as persistence times. According to Manousakis, the intervals are multiples of micro-event times of 5 or 10 ms. This is an obvious analogue to the relation between modular units $Q_0 \cong 4.5$ ms and $2Q_0 \cong 9$ ms and the respective coherence limits 30 × 4.5 ms = 135 ms and 30 × 9 ms = 270 ms. Furthermore, of two long-term characteristics, one, referred to as T_b and ranging from 0.4. to 0.9 seconds, is of the order of magnitude required for the generation of micro cycles (Geissler, 1991), the second, referred to as T(s) and ranging from 3 to 6.2 seconds and more, is of the size of maximum "shells" (Geissler, 1992), ensuring access to information on the sensory level. Thus the least one can say is that available evidence in BR conforms with the logic successfully applied in designing our experiment and extracting regularities from the data obtained.

A challenge to physiology

Considered as a general program the greatest potential of TQM may be seen in the fact that the structural and numerical invariants on which it builds are neutral against the choice of theoretical language and data basis from the realms of psychology or physiology. This provides the unique chance of convergent modeling based upon canonic parameterization. However, one should not underestimate the methodological difficulties involved in such an endeavor: TQM invariants relate to the necessary and sufficient basis of mental activity which Fechner once termed "the Psychophysical Process" (PP). Clearly, the machinery below and behind PP, necessary for its implementation, is beyond the scope of psychophysical experimentation and theorizing. More difficult to accept seems to be that substrate-oriented research in mental activity is reliant on the aid of psychophysical constructs when pinpointing and validating physiological equivalents of PP. The rationale of TQM is fairly radical in this respect. For example, while physiologically founded theories have rightly suggested that mental activity operates at the boundary of stability, TQM with the constant internal Weber Law according to H3 sets a quantitative norm for the transition from near-deterministic performance to chaotic activity, and it does so with the provisional claim of universal validity that can be refuted only by psychophysical means. Search for physiological equivalents of psychophysical constructs, if accepted as a reasonable goal, puts a heavy burden onto physiology, because substrate-oriented research is confronted with a degree of signal variability in general much larger than that found in psychophysical experiments. Therefore, dissociation of the relevant fractions of variability and even more so the identification of those properties that are invariant across the brain seem impossible without a psychophysical target structure and carefully controlled backward-search starting from psychophysical information. Mutatis mutandis, the same applies to our suggestion of a delay-based timing principle for which new massive complementary evidence has been presented in this article. It appears that corresponding mechanisms have so far generally not even been considered because the relevant transmission delays are inconspicuous by comparison with recordable oscillatory brain potentials. Demonstration of their presence and their relevance as a basis of timing mechanisms in different areas of the brain is a difficult aim certainly not to be pursued without related psychophysical evidence and at least rough estimates of their quantitative characteristics. But here they are.

Appendix

Integers $N \leq M = 30$ fall apart into the four subsets (in descending order) S1: {30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16}; S2: {15, 14, 13, 12, 11, 10, 9, 8}; S3: {7, 6, 5, 4}; S4: {3, 2} plus unity such that no element of one of these sets is divisor of another element of the same set. With the ideal estimate $Q_0 = 4.57$ ms, one obtains as corresponding sub-ranges of R_1 the intervals: R_{11}: $[30Q_0, 16Q_0]$, R_{12}: $[15Q_0, 8Q_0]$ and R_{13}: $[7Q_0, 4Q_0]$ and R_{14}: $[3Q_0, 2Q_0]$. Expressed in frequencies (in Hz) this corresponds to: R_{11}: $[7.3 \leq f \leq 13.7]$; R_{12}: $[14.6 \leq f \leq 27.4]$; R_{13}: $[31.9 \leq f \leq 54.5]$; and R_{14}: $[72.9 \leq f \leq 109.4]$. Of those, R_{11}, R_{12}, and R_{13} agree surprisingly well with standard definitions of the EEG alpha, beta and gamma bands, respectively. From the property of R_{11} as the set of intermittencies from which all others can be generated follows the hypothesis of a privileged position of the alpha band. Figure 6 illustrates how good the predicted band width agrees with employed alpha-band definitions.

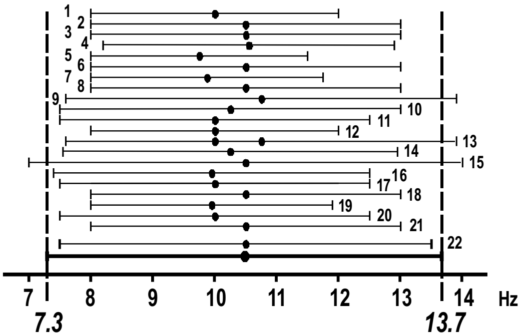


Fig. 6. TQM prediction for R_{11} (bold bar, dashed vertical line) as compared to alpha band definitions from the literature (1-21: after Klimesch et al. (2007),22: after Livanov (1971).

Acknowledgments

This contribution is based on an international cooperation launched in 2006 during a visit by H.-G. Geissler at Vilnius University and sponsored by grant 436 LIT 18/1/06 from the German Research Council DFG. The reported experimental research was funded by grant MIP-015/2011 from the Research Council of Lithuania. The draft of this paper was written and finalized by H.-G. Geissler. He takes responsibility also for details of theoretical interpretation. We thank Algirdas Bastys for technical assistance and Hanna Geissler for communication and help with the language. The first author wishes to thank Lawrence Ward for a stimulating chat.

References

Atmanspacher, H., Filk, T., & Römer, H. (2004), Quantum Zeno features of bistable perception, *Biological Cybernetics*, *90*, 33–40.

Békésy, G. von (1936). On thresholds in hearing and feeling of sinusoidal low-frequency air pressure oscillations (in German). *Annalen der Physik, 26*, 554-566.

Berger, (1929). Über das Elektroenkephalogramm des Menschen [On the human electroencephalogram]. *Archiv für Psychiatrie und Nervenkrankheite , 87*, 527-570.

Abeles, M., Vaadia, E., Bergman, H., Prut, Y., Haalman, I., & Slovin, H. (1993). Dynamics of neural interactions in the frontal cortex of behaving monkeys. *Concepts in Neurosciences, 4* (2), 131-158.

Blake, R. (2005). Landmarks in the history of binocular rivalry. In: D. Alais & R. Blake (Eds.), *Binocular Rivalry and Perceptual Ambiguity,* 1-28. Cambridge: MIT Press.

Bredenkamp, J. (1993). Relatedness of invariance hypotheses in the psychology of memory (in German). *Zeitschift für experimentelle und angewandte Psychologie*, 40, 368-385.

Burle, B., & Bonnet, M. (1999). What's an internal clock for? From temporal information processing to temporal processing of information. *Behavioural Processes*, 45, 59-72.

Burle, B., & Bonnet, M. (2000). High-speed memory scanning: a behavioral argument for a serial oscillatory model. *Cognitive Brain Research*, 9, 327-337.

Burle, B., & Bonnet, M. (2003). Behavioral and electrophysiological oscillations in information processing: A tentative synthesis. *Time and Mind II*. Ch. 12, 209-232. Göttingen: Hogrefe.

Eckhorn, R. et al. (1988) Coherent oscillations: a mechanism of feature linking in the visual cortex? Multiple electrode and correlation analyses in the cat. *Biol. Cybern., 60*, 121–130.

Engel, G. R. (1970). An investigation of visual responses to brief stereoscopic stimuli. *Quart. J. of Experimental Psychology,* 22, 148-166 (cit. O'Shea & Crassini, 1984).

Fomin S.V., Sokolov E.N., & Vaitkevicius, H.H. (1979). *Artificial sense organs* (in Russian). Moscow: Nauka. Quoted on page 178.

Fechner, G. (1860). Elemente der Psychophysik. Leipzig: Breitkopf und Härtel

Fries, P., Nicolic, D., & Singer, W. (2007). The Gamma Cycle. *Trends in Neurosciences, 30,* 7, 309-316.

Geissler, H.-G.(1985). Sources of seeming redundancy in temporally quantized information processing. In G. d'Ydewalle (Ed.), *Cognitive information processing and motivation.* 23rd International Congress of Psychology*, 3*, 119-228. Amsterdam: North-Holland.

Geissler, H.-G. (1987). The temporal architecture of central information processing: Evidence for a tentative time-quantum model. *Psychological Research*, 49, 99-106.

Geissler, H.-G. (1991). Time-code constants – a link between psychology and physiology for cognitive research? Hypotheses and deliberations on quantal structures in the alpha activity of the brain (in German). *Zeitschrift für Psychologie 199*, 2, 121-143.

Geissler, H.-G. (1992). New magic numbers in mental activity: On a taxonomic system for critical time periods. In Geissler, H.-G., Link, S. W. and J. T. Townsend (Eds.): *Cognition, Information Processing and Psychophysics*, 293-321. Hillsdale: Erlbaum

Geissler, H.-G. (1994). On key options of discrete temporal structuring in perception and cognition: Deliberations in matters of "time quanta" (in German). In Dörner, D. und E. van der Meer (Eds.): *Das Gedächtnis*, *19-52*. Göttingen: Hogrefe

Geissler, H.-G. (2009). On the value of constancy, the constancy of values, and the singular role of time in fundamental psychophysical laws. In Elliott, M.A. and co-editors, *Proceedings of the 25th Annual Meeting of the ISP*, 37-58. Galway: Snap Printing.

Geissler, H.-G., Schebera, F.-U., & Kompass, R. (1999). Ultra-precise quantal timing: Evidence from simultaneity thresholds in long-range apparent movement. *Perception & Psychophysics*, 61 (4), 707-726.

Geissler, H.-G., and Kompass, R. (2003). Invariants in Mental Timing: From Taxonomic Relations to Task-related Modeling. In Helfrich, H. (Ed.). *Time and Mind II*. Chapter 11, 185-207. Göttingen: Hogrefe.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review, 84* (3), 279-325.

Gray, C.M. et al. (1989) Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature 338*, 334–337.

Grossberg, S. (2000). The complementary brain:Unifying brain dynamics and modularity. *Trends in Cognitive Sciences, 4*, 233-246.

Klimesch, W., Sauseng, P., Hanslmayr, S., Gruber, W., & Freunberger, R. (2007). EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Research Review, 53,*63-88.

Klimesch, W. (2011). Evoked alpha and early access to the knowledge system: The P1 inhibition timing hypothesis. *Brain Research, 1408*, 52-71.

Kompass, R. (2004). Universal temporal structures in human information processing: A neural principle and psychophysical evidence. In Kaernbach, Chr., Schröger, E., & Müller, H.J. (Eds.), *Psychophysics beyond Sensation: Laws and Invariants of Human Cognition*, 451-479. Mahwah, NJ: Erlbaum.

Latour, P.L. (1967). Evidence of internal clocks in the human operator. *Acta Psychologica, 27,* 341-348.

Lisman, J.E., & Idiart, M.A.P. (1995). Storage of 7 ± 2 short-term memories in oscillatory subcycles. *Science*, *267*, 1512-1515

Livanov, M.N. (1972). *On the spatial organization of brain processes* (in Russian). Moscow: Nauka

Manousakis, E. (2009). Quantum formalism to describe binocular rivalry. *BioSystems, 98,* 57-66.

O'Shea, R.P., & Crassini, B. (1984). Binocular rivalry occurs without simultaneous presentation of rival stimuli. *Perception & Psychophysics, 36 (3),* 266-276.

Puffe, Martina (1990). Quantized speed-capacity relations in short-term memory. In Geissler, H.-G. (Ed.), *Psychophysical Explorations of Mental Structures*, 290-302. Göttingen/Toronto: Hogrefe & Huber.

Sherrington, C. S. (1904). On binocular flicker and the correlation of activity of corresponding retinal points. *British Journal of Psychology, 1*, 26-60.

Sompolinsky, H., & Tsodyks, M. (1994). Segmentation by a network of oscillators with stored memories. *Neural computation*, 6 (4), p.p. 642- 657

Teghtsoonian, R. (1971). On the exponents in Stevens' law and the constant in Ekman's law. *Psychological Review*, *78,* 71-80.

Teghtsoonian, R., & Teghtsoonian, M. (1997). Range of acceptable stimulus intensities: An estimator of dynamic range for intensive perceptual continua. *Perception & Psychophysics, 59 (5),* 721-728.

Treisman, M., Faulkner, A., Naish, P.L.N. & Brogan, D. (1990). The internal clock: Evidence for a temporal oscillator underlying time perception with some estimates of its characteristic frequency. *Perception*, 19, 705-743.

Vanagas, V., Balkelite, O., Bartusjavicus, E., & Kirvelis, D. (1976). The quantum character of the recognition process in the visual analysator of humans (in Russian). In Gleser, V.D. (Ed.), *Information Processing in the Visual System*, *2-30.* Leningrad: Academy of Sciences of the USSR.

Vanagas, V. (2001). Quantal timing as a consequence of the anticipatory activity of the nervous system, In Sommerfeld, Kompass & Lachmann (Eds.), Fechner Day 2001, *Proceedings of the 17th Annual Meeting of the ISP*, 647-652. Lengerich: Pabst Science Publishers.

†*This paper is dedicated to Michel Treisman, Boris Burle and Michel Bonnet for their pioneering contributions to our knowledge on timing in the brain.*

Psychophysical scaling in pain in newborns: component comparison analysis

Rosana Maria Tristão[1]
Naiara Viudes Martins Garcia[1]
José Alfredo Lacerda de Jesus[1]
Carlos Tomaz[2]

Faculty of Medicine[1] and Institute of Biology[2], Campus Darcy Ribeiro, Universidade de Brasília, DF, Brazil

Corresponding Author: RM Tristão (rmtt@unb.br)

Abstract

COMFORT behavior scale (Cb) has been applied in neonatal units to measure pain in newborns. This study aimed to analyze parameter´s scale testing if they differ in sensibility to pain. The scores were correlated to skin conductance activity (SCA). Thirty-six newborns were videotaped whilst heel prick. The images were analyzed in three intervals: before, during and after the pain event. Cb scale scores were compared to SCA variables: number of waves/sec and area under curves in three different time intervals after the pain event: 15, 30 and 180sec. All factors of Cb were sensitive to changes among periods during-before and after-during. Significant correlations values were found between Cb and number of waves ($r < 0.6$). Facial Tension was the gold standard response to pain meanwhile factors as Crying and Calmness can be considered poor indicators of pain. These results are discussed in terms of phenomenological approach and anxiety paradigm.

Keyword: COMFORT behavior scale, pain, neonates, skin conductance activity, anxiety.

Introduction

The evaluation and scaling of pain perception are fundamentally related to the clinical art dependent on patient's report, behavioral observation and physiological measures based on physical examination. Newborns follow these parameters except the ability to report what they perceive. Although the statement that newborns are able to perceive pain was controversial for many decades, nowadays there are enough evidence that they are capable of perceive and report pain both on their medical condition or clinical procedures performed as collection of blood, endotracheal suction, surgery or other invasive procedures (Harrison *et al.*, 2006). Pain is defined by the International Association for the Study of Pain (IASP) as "an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage" (Merskey H. Bogduk, 1994; IASP Task Force on Taxonomy, 1994, updated 2011). This definition has been often used in studies about pain despite theoretical models or kind of subject, but this worldwide use can´t prevent or hidden the truth of how difficult is to define pain, specially the pain perceived by non-verbal subjects. To assess pain in infants, behavioral and physiological parameters were designed to reduce the subjectivity linked to patient response and give more subsides for research in neonates who do not report the pain (Eriksson *et al*., 2008; Storm, 2008). Several pain scales have been introduced, either unidimensional or multidimensional which combines behavioral and physiological dimensions of pain (de Oliveira *et al.*, 2010) such as COMFORT scale (Ambuel *et al.,* 1992; van Dijk *et al.,* 2005), *Neonatal Infant Pain Scale* (Lawrence *et al.,* 1993), *Neonatal Facial Coding System* (Grunau & Craig, 1987) and *Neonatal Pain, Agitation and Sedation Scale* (Hummel *et al.,* 2008, Hummel *et al.,* 2010) among others.

Despite the crescent pool of validated pain scales for infants, the physiological parameters as heart rate (Pereira *et al.,* 1999, Padhye *et al.,* 2009, Jesus *et al.*, 2011), blood pressure, oxygen saturation (Hummel & van Dijk, 2006, Jesus *et al.*, 2011) have attended to assess pain and validate other instruments in newborns. Skin conductance activity (SCA) (Storm, 2000; Storm & Freming, 2002, Gjerstad *et al.,* 2008) also has been used as a measure of pain specially in monitoring context of infants under sedation or anesthetized (Eriksson *et al*., 2008; Storm, 2001; Hullett *et al*., 2009, Jesus *et al.*, 2011; de Oliveira *et al*., 2012). This type of sweat is dependent on the response of the cerebral cortex through the activation of the sympathetic nervous system (SNS) on the sweat glands, and regardless of weather or cardio respiratory conditions (Storm, 2001; Harrison *et al*., 2006; Storm, 2008).

Previous study has already compared skin conductance to the scores of the COMFORT scale and COMFORT behavior scale. Gjerstad, Wagner, Henrichsen & Storm (2008) analyzed skin conductance fluctuations during endotracheal suctioning and they founded better correlation with the increase in the COMFORT scale score than the variation of heart rate and arterial blood pressure. Considering that the skin conductance is the fastest measurement system of all, giving pain perception measurement in short intervals as 15 seconds, one possible question raised from these results is whether some dimensions of the COMFORT behavior scale are as sensitive and fast in revealing infant´s pain perception. If it is so, it can be used as alert signals to clinical teams in emergency contexts or limited contexts for use pain scales or the skin conductance device, a common situation in ICU's context. van Dijk adapted and modified this scale cutting off these physiological variables and created a COMFORT behaviour scale which uses only the behaviour and the phenomenological dimensions (van Djik *et al.,* 2005). This phenomenological dimension is linked to the behavioral one, and consists of at least two sub-dimensions, sensory-discriminative and affective-emotional (experience's spontaneous cognitive/conative reactions to own pain experiences). According to Aydede *et al.,* 2010 it means that "These reactions were conceived as forming *conative propositional attitudes*. In other words, the painfulness of pains was constituted by their power to immediately evoke in [one] the peremptory desire that the [pain] perception should cease". The sensory-discriminative aspect of pain is representational: it represents tissue damage. To define the phenomenological aspect is very difficult, but it can be said that feeling pain involves perception although perception doesn't exhaust its nature: feeling pain is also an affective/emotional experience that can be explained in terms of the functional role of pain's sensory/representational content.

Taking into account the above mentioned aspects, this study aims to compare all COMFORT behavior scale´s dimensions and its behavioral indicators to the physiological measurement of pain by skin conductance to test the sensibility and specificity of each scale´s item. Hence, this study assessed whether each variable in the COMFORT behavior scale can be used independently as a measure predictor of pain in newborns. To assess each variable of COMFORT behavior scale we used the validated equipment for measuring pain in neonates, Skin Conductance Measurement System (SCMS$^{®}$), as a comparative method of results analysis (Harrison, 2006) in three different time windows to verify the efficacy of each item of the scale. We hypothesized that the best correlations between behavioral and physiological indicators would reveal the most efficient responses of neonates to be used in future studies about acute pain in infants and the most effective response to be considered for clinical and research purposes.

Methods

Subjects

Were selected 36 newborns, 19 males and 17 females, with gestational age between 37 to 41 weeks (mean = 38,95 weeks; ± 1,35 weeks) with up to 48 hours of life at the Maternity of the University of Brasília Hospital. These babies were subjected to routine procedure in the ICU such as daily assessment of glucose by heel prick. Parents of infants were consulted and informed about the purpose of the study by signing the consent form. Were excluded from the sample infants with postnatal age less than 24 hours, with Apgar score less than seven in the fifth minute of life with a diagnosis of intracranial hemorrhage in the third or fourth degree (Volpe 2008); with metabolic, respiratory, circulatory, congenital disorders; and has used drugs that interfere with the perception of pain (analgesics, sedatives or muscle relaxants), neonates whose mothers used opioid and/or its derivatives during pregnancy.

Procedure, Apparatus and COMFORT Behavior Scale

The infants' responses were videotaped using a camcorder (DCR-SR 47, Sony), SCA was measured by the Skin Conductance Measurement System (SCMS®, Medstorm Innovation) through three established conductance variables: basal level of conductance peaks, number of waves per second (NWps) and area under the curve of waves (AUC). COMFORT behavior scale (C*b*) was used to assess the following variables: muscle tone, facial expression, alertness, calmness/agitation, respiratory pattern and physical movement. Each variable has a range of five points, obtaining a total sum of six behavioral variables. Thus, one can obtain six as lower value reflecting the minimum degree of discomfort and a maximum of 30 points (van Djik *et al.,* 2005).

To evaluate the behavioral response to noxious stimulus, the neonates were submitted to their daily routine heel prick for glucose monitoring. The exams were done early morning, at their bed and by the same examiner. The three electrodes of the equipment were attached and wrapped to the left foot 10 min before starting the observation period. The skin conductance measures were taken at three time intervals after the heel prick: 15, 30 and 180sec. All variables were analyzed in three periods "before" "during" and "after" the procedure. One trained observer assessed the movies and scored the infant´s behaviors by Cb. The SCA parameters were taken from the SCMS® software register system for three minutes.

Statistics

Each item's score of Cb was compared between the periods "before", "during" and "after" the procedure by the paired-related sample when comparing the differences *during-before*. The Wilcoxon nonparametric-paired test was used to analyze the significant difference of the change in scores along the studied periods. To assess the global score of Cb, as well as their variables, were compared to SCA by the Spearman's bivariate correlation. To verify if the clinical and demographic were related, it was used GLM two-way ANCOVA analyses. Analyses of agreement between Cb scores and subscores and skin conductance variables, all them calculated as difference between during-after pain periods, was made using Kendall's coefficient of concordance. Kolmogorov-Smirnov's normality test was passed for all data sets ($p > 0.05$) and the Levene's test of homogeneous variances was not significant for all analyses ($p > 0.05$). The data was stored on computer and analyzed using SPSS Package Version 17.0 and Minitab® 15.1.30.0.

Results

The sample variables had a mean gestational age of 38.9 weeks; 68,3% were delivered by cesarean section; 39% were large for gestational age; 41,5% small for date and 5,1% from diabetic mothers; 68% were breastfed one hour before the procedure. The paired-sample test of each variable of Cb showed significance for all variables both the period during–before and after–during (Table 1). All variables were statistically significance $p < 0,001$. According to the Spearman's bivariate correlation (Table 2) between each variable of Cb and SCA considering NWps after the procedure (15sec, 30sec and 180sec) all of them were significant ($p < 0.05$) but showed from fair to moderate correlations ($r < 0,6$). Crying and calmness were the weakest factors to all time windows (Table 2). Only physical movement and facial expression showed a significant correlation between AUC 15sec and 30sec after the procedure, but all of them showed from fair to moderate correlations ($r < 0,6$). Global score of the Cb was compared with the SCA by Spearman's Bivariate Correlation in each period of analysis. Were not found statistically significance between the score of the scale before and during the procedure (heel prick) with the NWps (15sec, 30sec, 180sec) and AUC (15sec, 30sec, 180sec). The Cb score after the heel prick was statistically significant ($p < 0,005$) for the NWps (15sec, 30sec, 180sec), but the highest correlation was to NWps 180sec ($r = 0,504$, $p < 0,001$).

Table 1. Paired-samples test of each variable of Cb in the 37 neonates.

Pair	During -before	After -during
Cb	-4,956**	-5,450**
Alertness	-4,057**	-3,849**
Calmness	-4,965**	-4,096**
Crying	-5,597**	-5,412**

Physical Movement	-4,818**	-4,757**
Muscle Tone	-4,954**	-4,789**
Facial Tension	-5,241**	-5,179**

Notes: Wilcoxon test ** $p < 0.01$;

Table 2. Spearman's Bivariate Correlation between each variable of Cb and SCA considering 15, 30 and 180 sec after the procedure (heel prick).

Variables	15sec	30sec	180sec
Cb and NWps	0,424**	0,383*	0,504**
Alertness and NWps	0,576**	0,468**	0,444**
Alertness and AUC	0,431**	-	-
Calmness and NWps	0,449**	0,405**	0,333*
Crying and NWps	0,328*	0,348*	0,437**
Physical Movement and NWps	0,560**	0,533**	0,609**
Physical Movement and AUC	0,424**	0,314*	-
Muscle Tone and NW	0,383*	0,404**	0,413**
Facial Tension and NW	0,418**	0,391*	-
Facial Tension and AUC	0,327*	-	-

Notes: * $p < 0.05$; ** $p < 0.01$

Additional analyses of agreement between Cb and skin conductance was made using Kendall's coefficient of concordance. The Kendall rank correlation coefficient was calculated to the "difference during-before pain event" values between Cb general score and factors´scores and SCA variables (NW and AUC). No Kendall's coefficient of concordance was found between general score of Cb and SCA (NW and AUC) at 15sec interval, to NW 30sec were found only marginal significant values , but to NWps 180sec interval, it was found high significant Kendall's tau-b value (.313, p=.006). Looking at the factors and the three time intervals of SCA variables, it was found to 15sec agreement between Facial Tension and NWps 15sec (Kendall's tau-b= -.303, p=.023). To NWps 180sec, again Facial Tension (Kendall's tau-b= .363, p=.001), and also Muscular Tonus **(**Kendall's tau-b= .270, p=.021), Physical Movement (Kendall's tau-b= .299, p=.006), and Alertness (Kendall's tau-b= .321, p=.002).

Discussion

In our study, almost every factor of Cb were statistically significant ($p < 0,005$) when correlated with the SCA, evaluating the NWps, in the periods of analysis of 15sec, 30sec and 180sec. Despite the statistical significance between the variables of the Cb, the overall score of the scale when compared with the SCA variables by Spearman's Bivariate correlation showed a weak correlation ($r < 0,50$). This result may suggest that the scale, although validated for the measurement of pain in newborns, behaves more for chronic pain, when the analysis time is longer than 2 minutes while SCA is more sensitive and specific for the measurement of acute pain in small instant of seconds (15sec) after the heel prick. Corroborating this idea, we found that there was a statistical correlation only when sub items of the scale were compared with the number of waves in180sec suggesting that the pain response is delayed even when taken into consideration the Cb. Only the variable of facial expression was not correlated with the NWps in 180sec. This can be due to the low carrying capacity of muscle seen in newborns or even due to misinterpretation of the observer. In fact, a study with 27 full-term healthy newborns, which compared in combination with behavioral and physiological measures of skin conductance, suggested that the AUC was the variable most sensitive and specific in measuring pain levels (Eriksson *et al.,* 2008). However, in this present study, correlation was observed during the 15sec with the factors analysis physical movement and facial expression with $p = 0.037$ $r = 0,327^{*}$ and $p = 0.006$ $r = 0,424^{**}$respectively. The lack of correlation with other variables of the Cb can be explained due to interference or artefact, making it necessary for automatic recognition of filtering.

The analyses of agreement between Cb and SCA by Kendall's coefficient of concordance pointed that its scores are more related to values of SCA more for late intervals (180sec), far from the time interval more related to the pain event (15sec). Facial Tension seems to be the gold standard of this scale as it kept in high levels all over the behavioural observation time. Otherwise, factors as Crying and Calmness can be considered poor indicators of pain. Other factors as Muscle Tone**,** Physical Movement and Alertness only agreed with time interval of 180sec. Nevertheless, all agreements were only to NWps and all were only in a fair level of agreement, which was expected as the SCA is a highly fast way to measurement compared to behavioral scales.

The Cb should be evaluated in structured its three dimensions or components that comprise (Ambuel *et al.*, 1992) physiological, phenomenological and behavioral. Analyzing under this theoretical approach, it is possible to assume that the SCMS® system, built to be sensitive not only to physiological dimension but also to the emotional one, and the facial tension factor of Cb are more related to the physiological level, while factors as alertness, physical movement and muscle tone are more related to the phenomenological dimension.

The phenomenological component of a behavioral scale, described as the personal account of pain, anxiety, among others, is the most difficult to assess in newborns since they are unable to speak. According to Ambuel *et al* (1992), anxiety can exist even in the absence of pain while the pain is related to an external noxious stimulus able to activate ascending cortical pathways for perception and interpretation of pain, and descending pathways to the development of a response (Anand *et al*., 2007). Associated with this mechanism, researchers assessed that these mechanisms are integrated with the sympathetic nervous system by changing the conductance of the skin due to the release of acetylcholine at postganglionic synapses of the muscarinic receptors of the sweat glands (Storm, 2008). According to the behavioral component, can be observed in this study that all the behavioral variables of the Cb when conducting an assessment using non-parametric pair before-during and during-after the procedure (heel prick) showed differences statistically significant but a weak correlation between SCA variables and the scale. Thus, it allows inferring that this variation in the score of each variable follows the variation of the scale of perception/response to pain by the newborn. Hence, the data also allow raising the hypothesis of responses like increase in physical movements, muscle tone and alertness is likely signals of anxiety than pain. To study this premise it could be suggested, in future studies, instead of considering the COMFORT´s entire interval of two minutes, to use varieties of observation behavior strategies as time sample, splitting the time window in smaller intervals searching for the decay of some responses and the raising of others. In that way it would be possible to define which responses are more related to the past painful event and which are related to fear of future pain.

Acknowledgments

This research was sponsored by grants from FAPDF, CNPq and FAHUB for the purchase of equipments and also received the support of the staff from Neonatal Intensive Care Unit, University of Brasilia Hospital, Brasília, Brazil. Special thanks to the mothers and babies. Thanks also to Marcos Vinícius Melo de Oliveira for the video analysis of the COMFORT scale.

References

Ambuel B., Hamlett K. W., Marx C. M. & Blumer J. L. (1992). Assessing distress in pediatric intensive care environments: the COMFORT scale. *J Pediatr Psychol, 17*, 95–109.

Aydede, M. (2010). "Pain", *The Stanford Encyclopedia of Philosophy (Spring 2010 Edition)*, Edward N. Zalta (Ed.) [http://plato.stanford.edu/archives/spr2010/entries/pain/; last accessed on 2012 jan 29].

de Oliveira M. V. M., de Jesus J. A. L. & Tristão R. M. (2012). Psychophysical parameters of a multidimensional pain scale in newborns. *Physiol. Meas., 33*, 39–49.

Eriksson M., Storm H., Fremming A. & Schollin J. (2008). Skin conductance compared to a combined behavioural and physiological pain measure in newborn infants. *Acta Pædiatrica, 97*(1), 27–30.

Gjerstad A. C., Storm H., Hagen R., Huiku M., Qvigstad E. & Raeder J. (2007). Comparison of skin conductance with entropy during intubation, tetanic stimulation and emergence from general anaesthesia. *Acta Anaesthesiol Scand, 51*(1), 8–15.

Gjerstad A. C., Wagner K., Henrichsen T. & Storm H. (2008). Skin conductance versus the modified COMFORT sedation score as a measure of discomfort in artificially ventilated children. *Pediatrics, 122*(4), 848-53.

Grunau R. V. & Craig K. D. (1987) Pain expression in neonates: facial action and cry. *Pain, 28*(3), 395-410.

Harrison D., Boyce S., Loughnan P., Dargaville P., Storm H. & Johnston L. (2006). Skin conductance as a measure of pain and stress in hospitalised infants. *Early Human Development, 82*(9), 603—608.

Hullet B., Chambers N., Preuss J., Zamudio I., Lange J., Pascoe E. & Ledowski T. (2009). Monitoring Electrical Skin Conductance: A Tool for the Assessment of Postoperative Pain in Children? *Anesthesiology, 111*(3), 513–7.

Hummel P., Puchalski M., Creech S. D. & Weiss M. G. (2008). Clinical reliability and validity of the N-PASS: neonatal pain, agitation and sedation scale with prolonged pain. *J. Perinatol., 28*(1), 55-60.

Hummel P., Lawlor-Klean P. & Weiss M. G. (2010). Validity and reliability of the N-PASS assessment tool with acute pain. *J. Perinatol., 30*(7), 474-8.

IASP Task Force on Taxonomy (1994). "Part III: Pain Terms, A Current List with Definitions and Notes on Usage" (pp 209-214) Classification of Chronic Pain, Second Edition, IASP Task Force on Taxonomy, edited by H. Merskey & N. Bogduk, IASP Press, Seattle.

Jesus, J. A. L.; Tristao, R. M.; Storm, H.; Rocha, A. & Campos Junior, D. (2011). Heart rate, oxygen saturation, and skin conductance: A comparison study of acute pain in Brazilian newborns. *Conference Proceedings IEEE Engineering in Medicine and Biology Society, 30*, 1875-1879.

Lawrence J., Alcock D., McGrath P., Kay J., MacMurray S. B. & Dulberg C. (1993). The development of a tool to assess neonatal pain. *Neonatal Netw., 12*(6), 59-66

Merskey H. & Bogduk N. (1994). Classification of Chronic Pain. Seattle: IASP Press: 210.

Padhye N. S., Williams A. L., Khattak A. Z. & Lasky R. E. (2009). Heart rate variability in response to pain stimulus in VLBW infants followed longitudinally during NICU stay. *Dev. Psychobiol.*, *51*(8), 638-49

Pereira A. L. S. T., Guinsburg R., Almeida M. F. B., Monteiro A. C., Santos A. M. N. & Kopelman B. I. (1999). Validity of behavioral and physiologic parameters for acute pain assessment of term newborn infants. *São Paulo Med. J.*, *117*(2), 72-80

Ranger M., Johnston C. & Anand K. J. S. (2007). Current Controversies Regarding Pain Assessment in Neonates. *Semin Perinatol, 31*(5), 283-288.

Storm H. (2001). The development of a software program for analyzing skin conductance changes in preterm infants. *Clinical Neurophysiology, 112*(8), 1562-1568

Storm H. (2008). Changes in skin conductance as a tool to monitor nociceptive stimulation and pain. *Current Opinion in Anaesthesiology, 21*, 796–804.

van Dijk M., de Boer J. B., Koot H. M., Tibboel D., Passchier J. & Duivenvoorden H. J. (2000). The reliability and validity of the COMFORT scale as a postoperative pain instrument in 0 to 3-year-old infants. *Pain*, *84*(2-3), 367-377.

van Dijk M., Peters W. B., van Deventer P. & Dick Tibboel (2005). The COMFORT Behavior Scale: A tool for assessing pain and sedation in infants. *Am J Nurs, 105*(1), 33-6.

Stevens, Riddell, Oberlander & Gibbins (2007). Assessment of pain in neonates and infants. In KJS Anand, BJ Stevens & PJ McGrath (Eds.), Pain in Neonates and infants (third Edition) (pp. 67-90).

Storm H. & Fremming A. (2002). Food intake and oral sucrose in preterms prior to heel prick. *Acta Paediat., 91*, 555-60.

van Dijk, M., Peters, J. W., van Deventer, P., Tibboel, D. (2005). The COMFORT behavior scale: a tool for assessing pain and sedation in infants. *Am. J. Nurs., 105*, 105-133

ARISTOTLE (c. 384-322 BC) AND SIZE-DISTANCE INVARIANCE

Helen E. Ross
Department of Psychology, University of Stirling FK9 4LA, Scotland
h.e.ross@stir.ac.uk

G. Loek J. Schönbeck
Anjeliersstraat 62 A-I, 1015 NH Amsterdam-Binnenstad, The Netherlands
loek@lucinus.eu

Abstract

In his book on Memory, Aristotle takes the comparison of the sizes of objects at different distances as an analogy of the comparison of time periods. He seems to be referring to the theory of size-distance invariance – that we take account of distance in a kind of geometrical manner when judging size. He proposed that there was a mental representation of the outside world, as distinct from a knowledge of distance brought by the travelling visual ray. He held to a concept of geometry that was developed from Pythagoras onwards and flourished later among the Stoics. This may be the earliest known reference to size-distance invariance.

Background to size-distance invariance

Size-distance invariance (SDI) is the idea that we perceive the linear size of objects by scaling the image size in proportion to the perceived distance. The ratio of an object's perceived linear size to its perceived distance is determined by the ratio of its real linear size to its real distance (which is equivalent to the object's real angular size). This is a geometrical construction of perception which has ancient origins. This view was supported by the Stoic philosophers, in opposition to the Epicureans. Epicurus (c. 341-270 BC) took the view that veridical perceptual knowledge was directly impressed on the sense organs by stimulus objects. He and his followers believed that vision was based on intromission – something entered the eye. The Stoics generally believed in extramission, or in a mixture of extramission and intromission – sensitive visual rays, or a cone of flux, went out from the eye and touched objects, and then brought back information (such as distance) to the eye. Influential Stoic writers who explicitly stated SDI were Posidonius (c. 135-51 BC), and Cleomedes (c. 2nd century AD) (Schönbeck, 1998; H.E. Ross, 2000). The astronomer, mathematician and optical theorist Ptolemy (c. 140 AD) also did so (*Optics* II, 56, Transl. Smith, 1996), and gave the diagram shown as Figure 1. The angle at E can represent the line AB at distance EB or GD at distance ED.

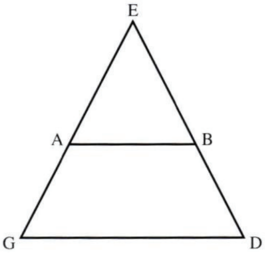


Figure 1. Ptolemy’s diagram for size-distance invariance

It is sometimes said that the geometrical account arose after the Epicurean account, because it was linked to the Stoics mentioned above. However, Euclid (c. 300 BC) was a contemporary of Epicurus and he discussed the role of geometry in calculating sizes on the basis of different distances. He drew on the work of previous mathematicians, such as Pythagoras (c. 570-495 BC) or even Thales (c. 600 BC). As a perception theorist, he stated that perceived size followed image size (angular size), but that linear size could be calculated from a knowledge of angular size and distance. He wrote in his *Optics* (Theorem 5): "Objects of equal size unequally distant appear unequal and the one lying nearer to the eye always appears larger" (Transl. Burton, 1945, p. 358). In this passage Euclid used the language of *appearances.* In another passage (Theorem 21 - To know how great is a given length') he argued that linear size (true object size) could be calculated in a geometrical manner from the angular size and the distance: but in this passage he used the language of *calculation* rather than appearances. This is not quite the same thing as SDI, which states that perceived size equals perceived linear size, and follows automatically from perceived distance and the true angular size. The difference between the two accounts depends on the level of consciousness at which the size scaling is said to take place (H.E. Ross, 2003).

The passage in Aristotle's *De Memoria et Reminiscentia*

There is a passage by Aristotle (c. 384-322 BC) which seems to imply that the concept of SDI was already around before the time of Euclid. It occurs in a book on Memory (*De Memoria et Reminiscentia - 452b7-22*), rather than in the more appropriate *De Anima* (On the Soul) or *De Sensu et Sensibilibus* (On the Senses and Perception). However, memory is closely allied to both sense perception and thought, because material for memory is presented to the mind through the "common sense" (whose organ is the heart) - where sensations from the five special senses are integrated (Murray, 1988). The passage's location in a section on memory may be why it has been overlooked by psychologists studying perception. We came across it by accident, because the translation included a diagram resembling SDI which caught our attention when looking for material on the moon illusion. This diagram (Figure 2) is reproduced from *De Memoria* (in *Parva Naturalia*, transl. J.L.Beare and G.R.T.Ross, 1931). The diagram was reconstructed from the Greek text by J.L. Beare and his editor W.D. Ross, working independently, according to Sorabji (1994) and Sisko (1997).

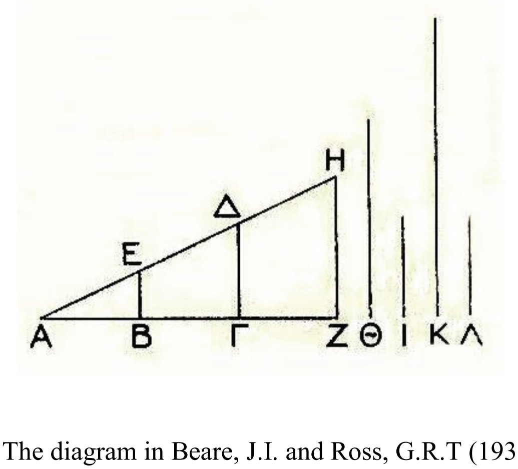


Figure 2. The diagram in Beare, J.I. and Ross, G.R.T (1931)

The text by Aristotle is ponderous and obscure, as are its translations. The various Greek manuscripts contain slightly different texts and alphabetic letters, and some contain diagrams. The translators choose a diagram to suit their interpretation of the text. Bloch (2007) thinks the manuscripts are too confused to give any diagram at all. Most recent translators give a diagram similar to Figure 2, with four lines shown beside the main triangle, but use the Roman alphabet instead of the Greek alphabet. All manuscript diagrams, and those of some of the earlier translators, show outer and inner triangles, corresponding to time and space experience respectively. This is illustrated in Figure 3, for the diagram of Freudenthal (1869). As Freudenthal (p. 417) explains, Aristotle sometimes used one letter to denote a whole line. In Figure 3 this is indicated by the curves H (running from M to H) and M (running from M to K). Freudenthal interprets BE as an inner representation of space, and KL as an inner represention of time. In this way he devises a kind of personal measuring stick which is independent of objective space and time. He may be hinting at SDI for spatial perception.

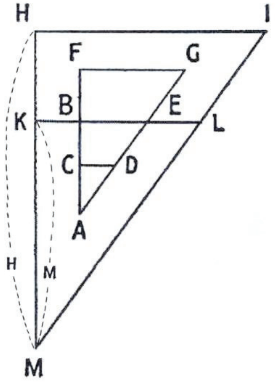


Figure 3. The diagram in Freudenthal (1869) with lettering by Hammond (1902)

We give our own translation below, to correspond to the Freudenthal diagram in Figure 3. The words in square brackets are added as our interpretation, but the material in round brackets is in the original text.

"But there is a most important fact to be noticed—that we must have apprehension of time either determinate or indeterminate. Let us grant as real something by which we discriminate greater and less periods. It is reasonable that we should do so in the same way as we discriminate extended magnitudes; we know things that have great size and are at a distance, not by our thought reaching out to them there, as some say our sight does (for non-existent things can equally be known), but by a proportional movement: for there exist in our thought figures and movements similar to the external objects. What then is the difference between thinking the objects of greater size and thinking the objects of smaller size? All the inner [representations] are smaller and the outer [objects] are proportionate; but it might well be that, just as in the case of the spatial things, the person has another corresponding representation within him, so it is with intervals. Thus, if one makes the movement AB, BE, this produces AC, CD (for AC and CD are in the identical ratio as AB and BE). What is it that makes AC, CD rather than AF, FG? Or, just as [with spatial objects] AF is to AB, thus [corresponding with time] H [MH] is to M [MK]. These processes, then, move together, but, if one wants to think [the external object] FG, one equally thinks [the corresponding representation] BE.Instead of the [time objective] ratio of H [MH] and I [HI] one thinks that inner [time representation] K [MK] to L [KL], for these [time objective MH and internal representation MK] are in the same proportion as [the object] FA stands to [the representation] BA."

Some authors give a diagram similar to that of G.R.T. Ross (1973, p. 115), which we show as Figure 4. The diagram is essentially the same as that of Beare and Ross (1931), apart from the difference in orientation and the use of the Roman alphabet. G.R.T. Ross (1973 p. 279) credits W.D. Ross with this diagram. He contends that Aristotle's aim was to show that external distances and movements are reproduced in miniature in the mental organ; the internal representations are proportional to the external realities, just as the sides of a small triangle are proportional to those of a much larger one

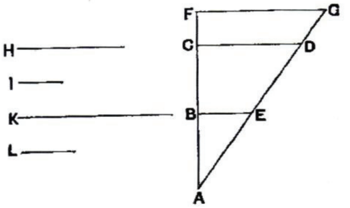


Figure 4. The diagram in Ross, G.R.T. (1973).

Roark (2011, p. 156-157) rejects the above interpretation. He argues that Aristotle was not concerned with the difference between external and internal magnitudes, but with the question of how differences in external magnitudes are represented. He maintains that, in Figure 4, the lines BE, CD and FG all have the same 'apparent' size because they subtend the same angle to the viewer at A; so how can the sizes be distinguished? Roark says that the key to answering this question is Aristotle's remark (in his translation) that "as we may assume within a person something proportionate to the forms of things, so we may equally well assume something else proportionate to their distances." (This is the part that we translate as "it might well be that, just as in the case of the spatial things, the person has another corresponding representation within him, so it is with intervals.") Roark then argues that Aristotle intended some kind of scaling of size according to distance, so that we perceive or imagine true object size rather than angular size. However, he makes no mention of the literature on SDI. He goes on to argue (p. 162-163) that this interpretation is consistent with some of Aristotle's other remarks about size. Aristotle says (*De Anima* 428b29-30) that our perception of the common sensibles is liable to error "especially when the perceived object is at a distance". For example, Aristotle says in at least three places that the sun appears to be one foot in diameter (*De Anima* 428b2-4; *De Insomniis* 458b28-9, 460b18-19). Aristotle wrote (*De Anima* 428b2-4) "But things also appear falsely, when one has at the same time a true supposition about them (e.g. the sun appears a foot across, but is believed to be bigger than the inhabited world)" (Transl. Schofield, 1978). Roark argues that we perceive the sun as small because we have no experience of seeing it at different distances, and it is vastly further away than any distances we see on earth. He adds that comparison with other distant celestial objects is of little help, and that the size of nearer occluding objects (e.g. mountains, clouds, the moon) gives only a minimal size to the sun. Thus, in the case of the sun, we do not have the ability to scale for distance. Again, Roark makes no mention of the large literature on size perception and the causes of the sun or moon illusion (e.g. H.E. Ross & Plug, 2002).

Sisko (1997) and Sorabji (2004. p.18) use the diagram of Beare and Ross shown in Figure 2. Sisko says that Aristotle probably held the view that the perception of magnitude by sight involves the production of small-scale analogues within the mind. Sorabji agrees that the diagram is concerned with judging the relative sizes of larger and smaller objects, and thinks that the mental diagram is more useful for calculating spatial distances than for calculating temporal ones (p. 108-109).

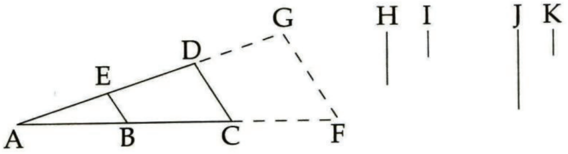


Figure 5. The diagram in Sachs (2004)

Sachs (2004, p. 177-178) gives a broadly similar translation and diagram to other recent authors. His diagram (Figure 5) is similar to that of D. Ross (1955, p. 250). He says that the appropriate diagram needs to be of the same kind as the one accompanying Proposition 12 in Book VI of Euclid's *Elements*. However, this proposition is about the geometry of proportional lines, and not about size and distance. There is of course a difficulty in saying how 'size' and 'distance' differ from 'length', and this remains a difficulty for modern versions of SDI (e.g. Schwartz, 1994). Aristotle's concept of 'movement' may well deal with 'distance' rather than 'length'. Sachs takes a different interpretation of the diagram from Roark, Sisko and Sorabji, saying "These lines are not meant to be representational images of any content in the imagination, but play the same role the lines do in the 'universal mathematics' of Book V of Euclid's *Elements*." Though he implies that the diagram is not about size scaling, he does say that it is similar to diagrams of Euclid. He points out that the 'something by which one distinguishes the time' is the common or primary perceiving power.

It is not obvious how the comparison of time periods is similar to the comparison of sizes and distances, though most of the commentators mentioned above discuss the issue. However, modern research shows that the ability to discriminate differences in time generally obeys Weber's law (Wearden and Lejeune, 2008).

Conclusions

The Aristotle manuscripts are indeed confused and hard to interpret. However, the interpretations of Freudenthal, Sisko, Sorabji and Roark are plausible, and they lend support to the idea that Aristotle is referring to the notion of SDI. In essence, Aristotle is saying that the comparison of periods of time is like the comparison of spatial sizes; we compare the latter in a geometrical manner, by taking account of distance, and keeping the ratios of sizes proportional to that of the distances. All of this takes place in 'thought'. In this passage Aristotle proposes a kind of mental gymnastics, as distinct from the work being done by an outgoing travelling ray (a view held by Plato and Empedocles – and elsewhere by Aristotle). He is thus quite modern in his idea that the mind (brain) creates a representation of the outside world. His description of spatial perception seems to be closer to that of perceptual SDI than to the mathematical calculations of Euclid. He does not say that perceived size is equivalent to mathematical angular size, but rather that it is equivalent to a range of linear sizes depending on the perceived distance. If this is a correct interpretation, it means that the idea of SDI was around much earlier than is usually thought.

References

Beare, J. I. & Ross, G.R.T. (Transl.) (1931). *Parva Naturalia*. In: *The Works of Aristotle* (ed. W.D. Ross). Oxford: Clarendon Press.

Bloch, D. (2007) *Aristotle on Memory and Recollection: Text, translation, interpretation, and reception in western scholasticism*. Leiden: Brill.

Burton, H.E. (1945) The Optics of Euclid. *Journal of the Optical Society of America,* 35, 357-372.

Freudenthal, J. (1869). Zur Kritik und Exegese von Aristoteles περὶ τῶν κοινῶν σώματος καὶ ψυχῆς ἔργων (parva naturalia). (Schluß von S. 81 ff.). II. Zu De memoria. *Rheinisches Museum für Philologie,* 24, 392¬419.

Hammond, W.A. (1902) *Aristotle's Psychology: A treatise on the principle of life (De Anima and Parva Naturalia).* London: Swan Sonnenschein & Co., pp. 208-210.

Murray, D.J. (1988) *A history of western psychology*. 2nd edition. Englewood Cliffs, N.J.: Prentice Hall.

Roark, T. (2011) *Aristotle on Time: A study of the physics*. Cambridge: Cambridge University Press.

Ross, D. (1955) *Aristotle: Parva Naturalia. A revised text with introduction and commentary.* Oxford: Clarendon Press.

Ross, G.R.T. (1973) *Aristotle: De Sensu and De Memoria*. New York: Arno Press.

Ross, H.E. (2000) Cleomedes (c. 1st century AD) on the celestial illusion, atmospheric enlargement, and size-distance invariance. *Perception*, 29, 863-871.

Ross, H.E. (2003) Levels of processing in the size-distance paradox. Ch.8 in Harris, L.R & Jenkin, M. (Eds.) *Levels of perception.* New York: Springer, pp.149-168.

Ross, H.E. & Plug, C. (2002) *The mystery of the moon illusion: Exploring size perception*. Oxford: Oxford University Press.

Sachs, J. (Transl.) (2004) *Aristotle's On the Soul and On Memory and Recollection*. Santa Fe, New Mexico: Green Lion Press.

Schofield, M. (1978) Aristotle on the imagination. In: Lloyd, G.E.R. & Owen, G.E.L. (Eds) *Aristotle on mind and the senses*. Proc. Seventh Symposium Aristotelicum. Cambridge: Cambridge University Press.

Schōnbeck, G.L.J. (1998) *Sunbowl or symbol. Models for the interpretation of Heraclitus' sun notion*. Amsterdam: Elixir Press.

Schwartz, R. (1994) *Vision: Variations on some Berkeleian themes*. Oxford: Blackwell.

Sisko, J.E. (1997) Space, time and phantasms in Aristotle, *De Memoria* 2, 452B7-25. *Classical Quarterly,* **47**(i) 167-175.

Smith. A.M. (1996) *Ptolemy's theory of visual perception: An English translation of the Optics with introduction and commentary*. Philadelphia: Transactions of the American Philosophical Society, Vol.86, Part 2.

Sorabji, R. (2004) *Aristotle on memory*. 2nd edition. Chicago: University of Chcago Press.

Wearden, J.H. & Lejeune, H. (2008) Scalar properties in human timing: Conformity and violations. *Quarterly Journal of Experimental Psychology*, **61**, 569-587.

AUDIOVISUAL SPEECH GAZE STRATEGIES

Willy Wong, Astrid Yi and Moshe Eizenman
Dept. of Electrical and Computer Engineering and
Institute of Biomaterials and Biomedical Engineering,
University of Toronto, Toronto, Canada
willy.wong@utoronto.ca

Abstract

The goal of this study was to examine the role of gaze in speech perception and to investigate gaze strategies for listening to speech in noise. Eye tracking was conducted on subjects engaged in a noisy audiovisual speech paradigm. Speech intelligibility was measured for eleven subjects listening to low-context sentences while viewing the talking face on a computer monitor. We found that speech intelligibility was similar for all fixations within 10° of the mouth area. However gaze strategy changed with speech signal-to-noise-ratio. When signal-to-noise-ratio was decreased, the number of gaze fixations to mouth region increased as expected. Other experiments were performed whereby gaze was fixed at different eccentricities and speech intelligibility was measured. These results were compared to results which would be obtained by mapping reduced acuity in the peripheral region to various levels of spatial degradation. Our findings suggest that the visual enhancement of speech occurs when subjects are able to see spatial frequencies of 6 cycles/degree or higher.

In this study we combined eye tracking with audiovisual speech to ask what salient visual information is required for speech perception in noise. Classic studies have shown that a talking face aids in speech intelligibility in noisy environments (Sumby and Pollack, 1954; Macleod and Summerfield, 1987). In our own studies, we have shown that auditory thresholds improves with co-modulated, synchronized visual signals (Luu, 2008; Qian. 2009). More generally, we know that the deaf are able to understand speech at a high level of proficiency on the basis of lipreading. An improved understanding of the role of vision in face-to-face communication can not only enhance our knowledge of how the brain processes speech, but also allows us to better find ways to develop engineering aids to help those with hearing impairment.

Despite the number of studies conducted on audiovisual speech, little is known about the role of vision in speech intelligibility. A study by Grant and Walden (1996) involving filtered speech showed that intelligibility was not affected when subjects were able to view the speaker's face. A similar study by Boothroyd et al (1998) showed that subjects had improved recognition performance in audiovisual speech when listening to speech encoded with only the fundamental frequency. Finally, the McGurk effect has spawned an entire area of study exploring how visual information (or visual information *mismatch*) plays a role in speech comprehension (McGurk and MacDonald, 1976).

Few theories have been proposed to explain how audiovisual enhancement of speech occurs. Most notable is the ‘common format’ theory which suggest that auditory and visual information transform to a common metric and are processed by cortical neurons responding to both audio and visual speech stimulations. (Calvert et al, 2000) Other studies have shown the co-dependence of the visual and auditory cortices. For example Pekkola et al has demonstrated that speechreading was found to activate the auditory cortex.

Our interest is in uncovering the patterns of gaze during speech perception in noise. Does there exist optimal gaze strategies and how does these strategies change with differing levels of noise? What happens to speech intelligibility when gaze is fixed at increasing viewing angles away from the face? Moreover, what is the role of peripheral vision in processing visual speech information and do experiments on spatial degradation reveal more about mechanism of audiovisual speech perception?

A number of studies on gaze behaviour and audiovisual speech perception have reported varying findings. One study showed that as auditory noise is increased, the number of fixations on the mouth increased. Other studies (Buchan et al., 2007), (Lansing & McConkie, 2003) have shown that a greater number of fixations were made on the nose and the mouth. Nevertheless, the general tendency of these studies suggest that the areas of primary fixation were on the eyes, nose and mouth. However, there hasn't been to our knowledge studies that examine the general role of gaze strategy within audiovisual speech perception.

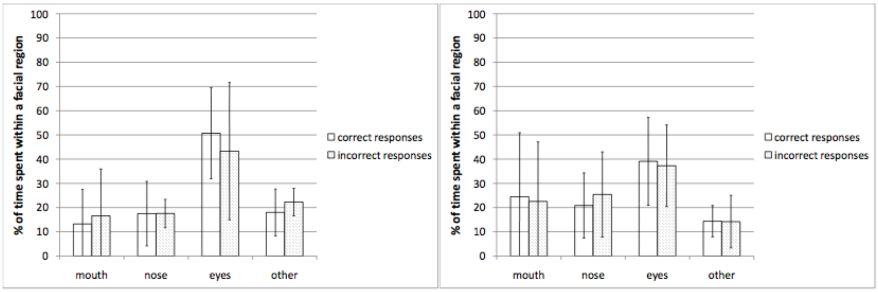
Method

Eleven naive adult subjects (1 female and 10 males) between the ages of 20 to 25 years old participated in the study consisting of 2 experimental sessions. All subjects were fluent in the English language with self-reported normal hearing and normal/corrected to normal visual acuity. The studies were approved by the Office of Research Ethics at the University of Toronto, and all subjects read and signed a consent form prior to the commencement of the research.

Low-context SPIN (Speech Perception in Noise) sentences were chosen from (Kalikow & Stevens, 1977) to minimize the probability that participants can determine the target word (last word) based on the initial words of the sentence. Sentences were spoken by a male talker fluent in the English language and recorded individually using a video camera. The audio stream (the speech signal) was normalized and combined with white noise to produce a new audio stream. The desired auditory SNR level was achieved by varying the amplitude of the speech signal level while maintaining the noise level.

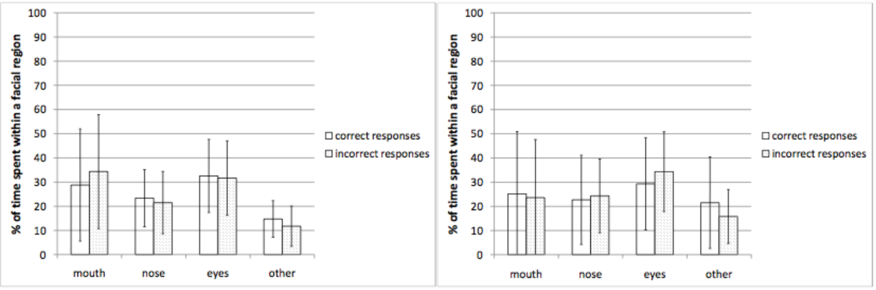
The experimental stimuli were presented on a 19" LCD monitor with audio output through headphones (AKG K301xtra). An advanced remote, non-contact point-of-gaze estimation system requiring only a single point subject calibration routine and consisting of 2 cameras and 4 infrared light sources (Guestrin & Eizenman, 2006) was used to monitor subjects' eye movements. The system provided a point-of-gaze estimation with accuracy of better than 1 degree at a sampling rate of 30 Hz. The experiments were carried out in a quiet laboratory room.

The experiments began with an estimate of the subject's audiovisual speech reception threshold. Within an experimental session, the presentation order of sentences and test conditions were randomized and no sentence was repeated. For the second experimental sessions, subjects were asked to maintain their gaze on a fixation cross, which was placed either at 0˚, 2.5˚, 5˚, 10˚, or 15˚ from the center of the mouth of the talker. The first 3 fixation points (0˚, 2.5˚, 5˚) were chosen to correspond to the primary fixation regions that were found in previous studies (Vatikiotis-Bateson et al., 1998; Buchan et al., 2007; Lansing & McConkie, 2003). These points were mapped to the mouth, nose, and eyes of the talker, respectively. The remaining fixation points (10˚, 15˚) were chosen such that subjects' speech intelligibility could be tested beyond the primary fixation regions. These points were mapped to the top of the hair of the talker and to the top of the computer monitor. All fixation points were vertically aligned with the center of the mouth of the talker.



(a) No Noise

(b) Audiovisual Threshold + 5 dB



(c) Audiovisual Threshold

(d) Audiovisual Threshold - 5 dB

Fig. 1. Average audiovisual speech intelligibility scores as a function of auditory SNR when subjects gazed naturally at talker.

Subjects were then presented with 4 sets of SPIN sentences which tested their speech intelligibility under 4 different audiovisual conditions: at threshold, threshold – 5 dB, threshold + 5 dB, and no noise. For each of these conditions, subjects were free to look anywhere on the computer monitor while reporting the last word heard after each stimulus presentation. In the second session, subjects completed 5 sets of trials while fixating on a cross that was placed at a specific distance from the center of the mouth (0˚, 2.5˚, 5˚, 10˚, and 15˚).

Results and Discussion

Natural Gaze as a Function of Noise and Time

In this experiment, we measured where a subject looked during a noisy speech perception task. Subjects were free to gaze anywhere during the task and the eye tracker recorded where the person looked. Fig. 1 shows data pooled across different subjects. No discernible pattern is observed for where a person looked as function SNR. The data for both correct and incorrect responses are shown. The tendency for the correct responses to mirror the incorrect responses demonstrates that an incorrect response is not likely to be due to the fact that the person was looking at the wrong region. In terms of the temporal variation of gaze (i.e. duration of stimulus), we reanalyzed the data to determine average gaze (as measured by Euclidean distance to mouth) as a function of time. The results, not shown here, show a clear monotonic convergence of gaze towards the mouth by the end of the sentence for all conditions, although convergence is stronger for noisier conditions (lower SNR).

Natural Gaze Strategies

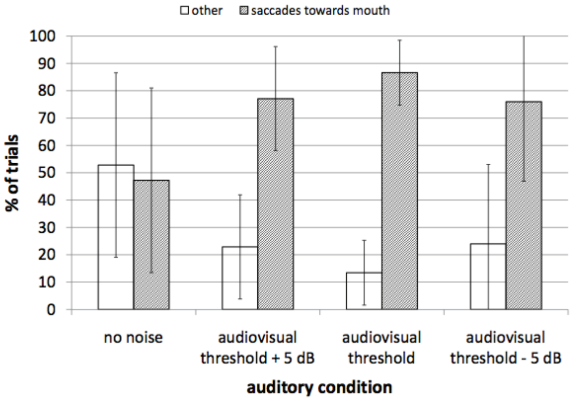


Fig. 2. Gaze strategies comparing the percentage of trials where there were saccades towards the mouth versus no saccades towards the mouth for different SNR conditions.

We also attempted to parse out different gaze strategies used by the subjects. We identified two major strategies. The first strategy entitled "saccades towards mouth" describes the strategy where subjects shifted their gaze from an initial starting point to a region within 2.5˚ of the mouth. 2.5˚ was selected because this encompassed 98% of our data for cases where subjects shifted their gaze to the mouth. The second strategy describes all other situations where the person did not fixate on a region within 2.5˚ of the mouth. The results we obtained (figure 2) show that the tendency is for subjects to gaze at the mouth when speech comprehension becomes difficult. While we would expect monotonic behaviour in terms of increased number of looks towards the mouth as SNR decreased, we could not draw such a conclusion from our data.

Audiovisual Performance Under Constrained Gaze

Very few experiments have been conducted to explore performance of audiovisual speech perception with respect to the proximity of fixation to visual cues. In this experiment, we carried out the same procedure but this time the subjects were asked to fixate on a cross placed either at the center of the mouth (0˚) or at 2.5˚, 5˚, 10˚, or 15˚ relative to the center of the mouth. The results in figure 3 illustrate that audiovisual performance is unchanged when the gaze is within 10˚ of the mouth. This is surprising in that this would indicate that much of the relevant visual information can be obtained by looking at just about anywhere on the face. The result also supports some very basic notions that we have about face-to-face communication -- that most people tend to look at the eyes when communicating with another person. According to these findings, however, looking at the eyes does not imply that the person will then "miss out" on the visual cues of audiovisual speech. They can still get this information provided that they are looking within 10˚ of the mouth.

The Role of Peripheral Vision in Audiovisual Speech

In the periphery, the ability to resolve fine spatial details is limited. Past studies have suggested that peripheral vision may be sufficient for audiovisual speech perception (e.g. Munhall et al, 2004). However, no study has explicitly investigated the role of peripheral vision in audiovisual speech perception, nor has there been a study which compared audiovisual speech perception between the conditions of viewing spatially filtered images with *foveal/parafoveal vision* and viewing unfiltered information with *peripheral vision*.

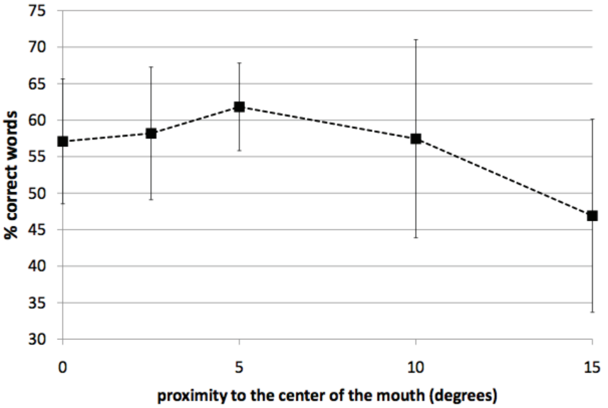


Fig. 3. Recognition performance as a function of fixed proximity of gaze from mouth.

Using a grating visual acuity curve, eccentricities were mapped to levels of spatial degradation. We then compared speech intelligibility performance by using this mapping. The performance was found to be identical (see figure 4). Our findings suggest that when subjects viewed low-pass filtered video recordings, their speech intelligibility was optimal (in terms of visual enhancement of speech perception) so long as they were able to see spatial frequencies below 6 cycles/degree.

Acknowledgements

This research was supported by the National Sciences and Engineering Research Council of Canada. We would also like to thank Eric Dacquay for his assistance with the experiments and with various technical matters in the project.

References

Boothroyd, A., Hnath-Chisolm, T., Hanin, L., & Kishon-Rabin, L. (1988, Dec.). Voice fundamental frequency as an auditory supplement to the speechreading of sentences. Ear and Hearing, 9(6), 316-312.

Buchan, J. N., Paré, M., & Munhall, K. G. (2007, Mar.). Spatial statistics of gaze fixations during dynamic face processing. Social Neuroscience, 2(1), 1-13.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000, Sept.). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Current Biology, 10, 649-657.

Grant, K. W., & Walden, B. E. (1996, Oct.). Evaluating the articulation index for auditory-visual consonant recognition. Journal of the Acoustical Society of America, 100(4), 2415-2424.

Guestrin, E. D., & Eizenman, M. (2006, Jun.). General theory of remote gaze estimation using the pupil center and corneal reflections. IEEE Transactions on Biomedical Engineering, 53(6), 1124-1133.

Lansing, C. R., & McConkie, G. W. (2003, May). Word identification and eye fixation locations in visual and visual-plus-auditory presentations of spoken sentences. Perception & Psychophysics, 65(4), 536-552.

Luu, S. (2008). Visual enhancement of auditory detection in the quiet. Unpublished master's thesis, University of Toronto.

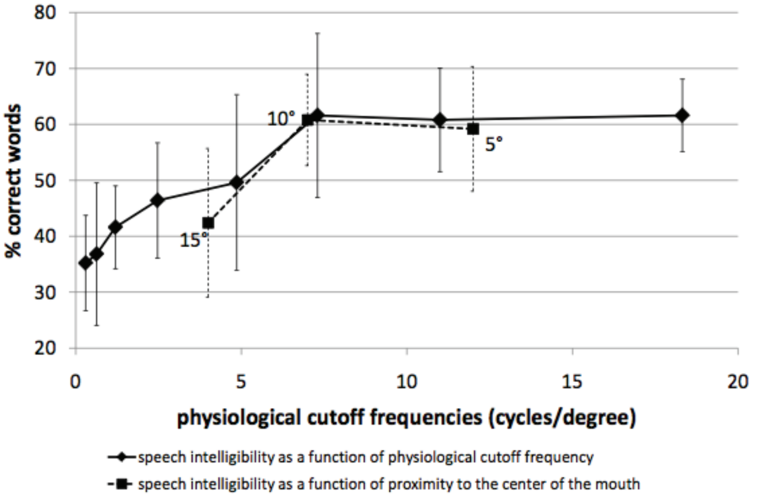


Fig. 4. Comparison of results between spatially-filtered foveal vision versus unfiltered peripheral vision.

McGurk, H., & MacDonald, J. (1976, Dec). Hearing lips and seeing voices. Nature, 264, 746-748.

MacLeod, A., & Summerfield, Q. (1987, May). Quantifying the contribution of vision to speech perception in noise. Journal of Audiology, 21(2), 131-141.

Munhall, K. G., Kroos, C., Jozan, G., & Vatikiotis-Bateson, E. (2004, May). Spatial frequency requirements for audiovisual speech perception. Perception & Psychophysics, 66(4), 574-583.

Pekkola, J., Ojanen, V., Autti, T., Jaaskelainen, I. P., Mottonen, R., Tarkiainen, A., et al. (2005, Jan.). Primary auditory cortex activation by visual speech: an fMRI study at 3T. Neuroreport, 16(2), 125-128.

Qian, C. L. (2009). Crossmodal modulation as a basis for visual enhancement of auditory performance. Unpublished master’s thesis, University of Toronto.

Sumby, W. H., & Pollack, I. (1954, Mar.). Visual contribution to speech intelligibility in noise. Journal of Acoustical Society of America, 26(2), 212-215.

Vatikiotis-Bateson, E., Eigsti, I.-M., Yano, S., & Munhall, K. G. (1998, Aug.). Eye movement of perceivers during audiovisual speech perception. Perception & Psychophysics, 60(6), 926-940.

PERCEIVING FILLED VS. EMPTY TIME INTERVALS: A COMPARISON OF ADJUSTMENT AND MAGNITUDE ESTIMATION METHODS

Emi Hasuo[*], Yoshitaka Nakajima[**], Erika Tomimatsu[**], Simon Grondin[*], Kazuo Ueda[**]
[*]Université Laval, Québec, Canada
[**]Kyushu University, Fukuoka, Japan
emi.hasuo.1@ulaval.ca

Abstract

A time interval between the onset and the offset of a continuous sound (filled interval) is often perceived to be longer than a time interval between two successive brief sounds (empty interval) of the same physical duration. The present study examined the occurrence of such phenomenon, sometimes called the filled duration illusion, for time intervals of 40-520 ms with the method of adjustment and the method of magnitude estimation. When the method of adjustment was used, the filled duration illusion appeared clearly for a few participants, while it did not appear for the majority of participants. With magnitude estimation, the filled duration illusion was more likely to occur. The amounts of the illusion did not correlate between the two methods, and it was suggested that even for the same participant, the perception of the empty and the filled intervals can be influenced by the experimental methods.

The perception of a time interval can be influenced by the structure of the interval, i.e., the way it is marked (e.g., Grondin, 2010). The interval is said to be *filled* when there is one continuous signal, and its onset and offset mark the beginning and the end of the time interval. The interval is said to be *empty* when it is marked by two successive brief signals and contain no stimulation within the interval. Previous studies showed that a time interval is perceived to be longer when it is filled than when it is empty (e.g., Wearden et al., 2007). Such phenomenon is sometimes called the *filled duration illusion*, and it has been reported repeatedly with various stimulus patterns and experimental tasks (Craig, 1973; Wearden et al., 2007, Zwicker, 1969/70).

However, a recent study by Hasuo et al. (2011) showed that the filled duration illusion does not take place in some cases. When subjective durations of very short time intervals of 20-180 ms were measured using the method of adjustment, the filled duration illusion occurred for some participants, while the filled intervals were perceived to be shorter than empty intervals—an opposite effect occurred—for more than half of the participants. The participants in which the filled duration illusion occurred and those in which it didn't were divided clearly by a cluster analysis. These findings of Hasuo et al. (2011) suggested that even for the same stimulus, the perceived duration could differ clearly between participants.

In the present study, we conducted two experiments to examine the robustness of Hasuo et al.'s (2011) results with longer durations (up to 520 ms) and with a different experimental task. In Experiment 1, the method of adjustment, as in Hasuo et al. (2011), was employed, and in Experient 2, the method of magnitude estimation. The stimuli in both experiments were the same, and some of the participants took part in both experiments, thus the results of the two experiments could be compared.

Experiment 1

The aim of Experiment 1 was to examine the occurrence of the filled duration illusion in a paradigm slightly modified from Hasuo et al. (2011). The experiment consisted of two sub-experiments, Experiment 1A and Experiment 1B.

Method

Participants Thirty-six undergraduate students of Department of Acoustic Design, Kyushu University, participated for course credits. Nineteen participants were assigned to take part in Experiment 1A, and the remaining seventeen to Experiment 1B.

Stimuli All sounds were 1000-Hz pure tones with a rise and a fall time of 10 ms. Empty intervals were marked by the onsets of two 20-ms sounds, and filled intervals were marked by the onset and the offset of a continuous sound (Figure 1). The presentation level of a 20-ms sound was 71 dBA, measured as the level of a continuous tone of the same amplitude, and the total energy of a filled-interval sound was equal to that of two 20-ms sounds together.

Each presentation consisted of a standard and a comparison in this order. The standard began 2.0-2.5 s after the participant clicked the "play" button on the computer screen, and the comparison began 2.5-3.0 s after the standard ended. In Experiment 1A, the comparison was a filled interval, and in Experiment 1B, an empty interval. The standard was the same in Experiment 1A and 1B: both empty and filled intervals were used as the standard.

The standard duration was 40, 100, 160, 280, 400, or 520 ms. Thus, there were 12 experimental conditions (2 [interval types: empty/filled] × 6 [standard durations]), both in Experiment 1A and in Experiment 1B.

Procedure The task for the participants was to adjust the duration of the comparison, by clicking the buttons and sliding bars presented on the computer screen, to make it subjectively equal to that of the standard. The participant could listen to the stimulus pattern and adjust the comparison duration as many times as he/she wanted, and the final comparison duration in each trial was recorded as the *point of subjective equality, PSE*.

For each condition, there were an ascending series and a descending series, and the PSEs from these series were averaged for each participant. Before these trials, 6 practice trials were carried out. The whole experiment, including the practice trials, took about 35 minutes per participant.

Results

Figures 2a and 2e show the mean PSEs of all participants in Experiments 1A and 1B. In both cases, the PSEs of the filled intervals (closed circles) were not so large compared to those of the empty intervals (open squares), i.e. the filled duration illusion did not appear. The standard deviations between participants (shown with error bars) were larger for empty intervals in Experiment 1A (Figure 2a), and for filled intervals in Experiment 1B (Figure 2e). This could

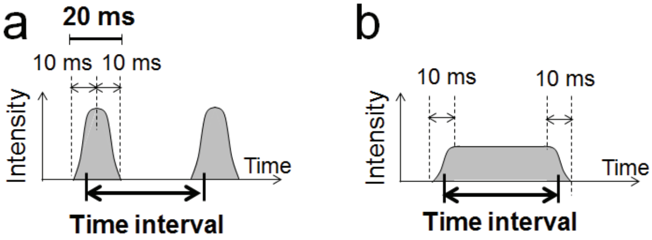


Figure 1. Illustration of the empty (a) and the filled (b) intervals. Note that the temporal midpoints (or beginnings depending on how we describe the patterns) of the rise/fall time were considered as the beginning and the end of a time interval.

have been due to the difficulty to compare the durations of an empty interval and a filled interval directly, which was necessary in these conditions.

To look more closely into these large variabilities, we conducted a hierarchical cluster analysis for standard durations of 100-520 ms[1], as in Hasuo et al. (2011), using the amount of overestimation of the empty intervals [(empty PSE) – (filled PSE)] for Experiment 1A and of the filled intervals [(filled PSE) – (empty PSE)] for Experiment 1B. Note that the filled-interval condition served as the control for Experiment 1A, and the empty-interval condition for Experiment 1B, since the comparison stimulus was the same as the standard. Clusters were determined by the Ward method, which analyzed the squared Euclidean distance between points.

Results of the cluster analysis showed that participants could be divided clearly into two groups (Figures 2d and 2h). We calculated the mean PSEs for Cluster 1 (16 out of 19 participants in Experiment 1A; 13 out of 17 participants in 1B) and Cluster 2 (3 out of 19 participants in Experiment 1A; 4 out of 17 participants in 1B) separately, and plotted them against standard duration (Figures 2b, 2c, 2f, 2g). For participants in Cluster 1, the mean PSEs did not differ much between the filled and the empty intervals (Figures 2b and 2f), whereas participants in Cluster 2 under/overestimated the empty/filled intervals, especially as the standard duration became longer (Figures 2c and 2g).

Summarizing Experiment 1, the filled duration illusion did not take place for many participants (Figures 2b and 2f), however, it took place clearly for a few participants (Figures 2c and 2g). This was in line with Hasuo et al. (2011). Extending the range of duration up to 520 ms did not seem to have much impact on the results.

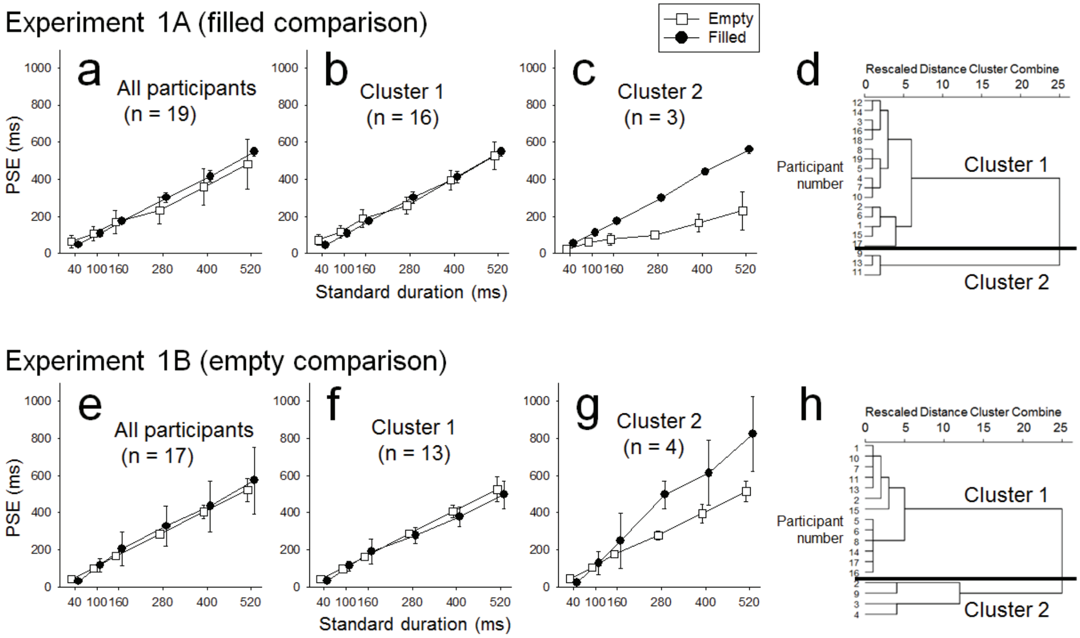


Figure 2. Results of Experiments 1A (a-d) and 1B (e-h): mean PSEs of all participants (a, e), Cluster 1 participants (b, f), and Cluster 2 participants (c, g). Clusters were divided according to the dendrograms (d, h) established by the hierarchical cluster analyses. Error bars represent the standard deviation between participants.

---

[1] We excluded the 40-ms condition from analysis because, when the standard was 40 ms, there were a few participants who tried to make the comparison shorter than the shortest duration technically possible (i.e., 10 ms in Experiment 1A, and 20 ms in Experiment 1B). The 40-ms condition was excluded from the analysis of Experiment 2 also, because the main purpose of this analysis was to compare the results with those of Experiment 1.

Experiment 2

The aim of Experiment 2 was to examine the occurrence of the filled duration illusion with the same stimuli as in Experiment 1 but with a different method, i.e., magnitude estimation.

Method

Participants Seventy-six undergraduate students of Department of Acoustic Design, Kyushu University, participated for course credits. Thirty-six had participated in Experiment 1.

Stimuli The empty and the filled intervals were the same as in Experiment 1 (Figure 1). At the beginning of each trial, the type (i.e. filled or empty) of the interval to be presented was indicated on the computer screen, and 2 seconds later, the time interval was presented once. Following the time interval, there was a silent period where nothing happened for 6 seconds, and then the next trial began automatically with the indication of the interval type for the next interval.

The interval duration was 40, 100, 160, 280, 400, or 520 ms. Thus, there were 12 experimental conditions (2 [interval types: filled/empty] × 6 [interval durations]).

Procedure The task for the participants was to listen to the time interval and verbally respond during the following silence the value that corresponded to the perceived duration of the interval. The participant could respond in any range of values he/she liked (including decimals and fractions) as long as they were positive numbers.

The 12 stimuli were presented in random order, and there were 4 repetitions. For each participant, the response value for each condition was obtained by calculating the geometric mean of the last 2 repetitions. Before these trials, participants listened to all 12 stimuli once, presented also in random order, without making responses. The whole experiment, including the listening-only trials, took about 8 minutes per participant.

Results

Figure 3a shows the geometric mean of the responses obtained from 76 participants plotted against the interval duration. The values of the responses for the filled intervals were larger than those for empty intervals (i.e., the filled duration illusion appeared), and this difference increased as the interval duration became longer.

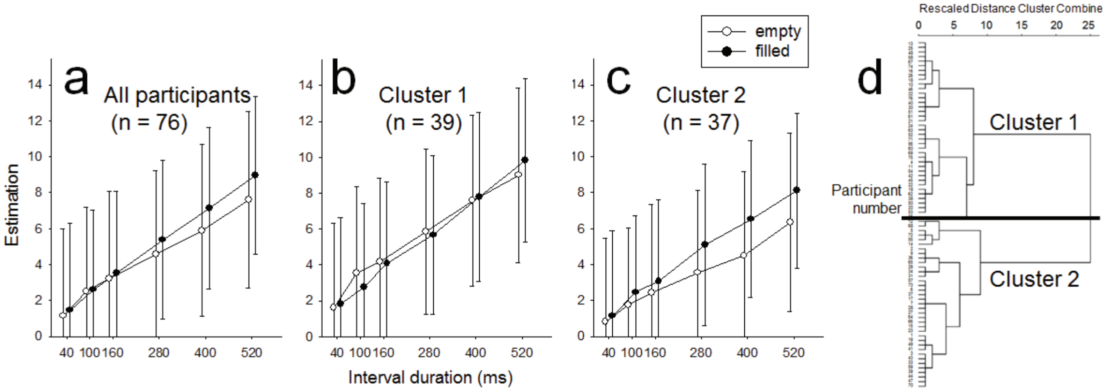


Figure 3. Results of Experiment 2: geometric mean magnitude estimations (responses) of all participants (a), Cluster 1 participants (b), and Cluster 2 participants (c). Clusters were divided according to the dendrogram (d) established by the hierarchical cluster analysis. Error bars represent the geometric standard deviation between participants. Note that the large variability is due to the huge difference in the range of response values chosen by each participant, and the error bars are shown just for the comparison between empty and filled intervals.

As in Experiment 1, we conducted a hierarchical cluster analysis using the log-transformed values of the amounts of overestimation of the filled interval [(filled geometric mean response) / (empty geometric mean response)] for interval durations of 100-520 ms[1]. Results showed that participants could be divided clearly into two groups (Figure 3d). Participants classified in Cluster 1 (39 out of 76 participants) did not show much difference between the filled and the empty intervals (Figure 3b), whereas participants in Cluster 2 (37 out of 76 participants) under/overestimated the empty/filled intervals, especially as the standard duration became longer (Figure 3c).

Summarizing, the filled duration illusion appeared clearly for some participants while it did not for other participants. This was similar to Experiment 1, but the filled duration illusion seemed more likely to appear with the method of magnitude estimation than with the method of adjustment (compare Figures 2 and 3).

General Discussion

The two experiments showed that, even for the same stimuli, the filled duration illusion can occur for some participants while it doesn't occur for others. This was consistent with Hasuo et al. (2011). To examine whether the participants who showed clear filled duration illusion in the method of adjustment task (Experiment 1) also showed clear illusion in the magnitude estimation task (Experiment 2), we focused on the 36 participants who participated in both Experiments 1 and 2, and calculated the amount of filled duration illusion for each of those participants for each experiment, by dividing the response for the filled intervals by the response for the empty interval (i.e., [(filled PSE) / (empty PSE)] for Experiment 1, and [(filled geometric mean response) / (empty geometric mean response)] for Experiment 2). Then, we averaged the amounts of filled duration illusion for the 100-520-ms intervals for each participant.

Figure 4 is the scatter plot showing the amount of filled duration illusion in Experiment 2 as a function of the amount of filled duration illusion in Experiment 1. It seemed that the participants who showed clear filled duration illusion in Experiment 1 did not always show large filled duration illusion in Experiment 2, and vise versa. Pearson's correlation coefficient (for 100-520-ms intervals) was low and non-significant (filled comparison, $r = .206$, $p = .396$; empty comparison, $r = .249$, $p = .336$)

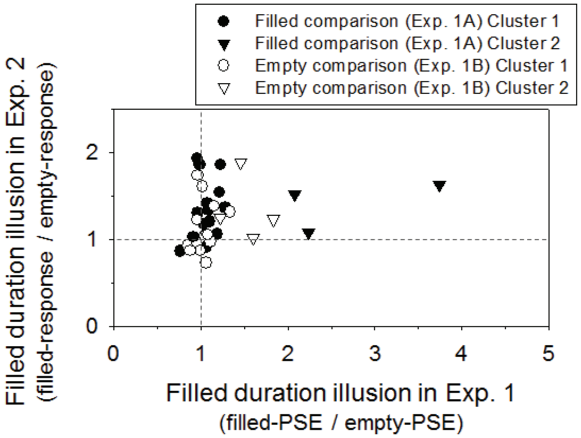


Figure 4. Scatter plot of the amount of filled duration illusion in Experiment 1 vs. Experiment 2 for the 36 participants who participated in both experiments. Larger values correspond to greater amount of filled duration illusion, and values smaller than 1 mean that the empty intervals were subjectively longer than the filled intervals (i.e., an opposite effect).

One interesting aspect about the results was that the participants were always divided clearly into two groups (Figures 2d, 2h, 3d), and that clear filled duration illusion appeared in one group whereas neither over- nor underestimation of the filled intervals appeared in the other. For participants with clear filled duration illusion, the amount of illusion increased as the interval lengthened (Figures 2c, 2g, 3c). This kind of filled duration illusion can be explained by assuming that the existence of a sound during a filled interval made the internal pacemaker run faster, which lead to the impression of time passing more quickly, consequently leading to an overestimation of filled intervals compared to empty intervals (Wearden et al., 2007). For participants without filled duration illusion (Figures 2b, 2f, 3b), such kind of pacemaker acceleration may not have occurred much; the subjective durations of the empty and the filled intervals were close to each other irrespective of the duration in their case. It could be possible that these participants were very accurate in perceiving the timing of sound onsets and offsets, thus the filled duration illusion did not appear (Repp & Marcus, 2010).

It was a new finding that there are two different modes for perceiving empty and filled intervals (one would cause filled duration illusion and the other would not), and that which mode the listener uses can be influenced by experimental methods. The filled duration illusion being less likely to occur for the method of adjustment task could be related to the fact that participants were able to listen to the stimulus pattern many times. This may have promoted the participant’s attitude to attend more carefully to the onset and offset. For the magnitude estimation task, having the participant make a judgment right after presenting the stimulus only once could have made the participants base their judgments more on the impression of the duration, not precisely on the onset and offset timing. This explanation warrants further investigation.

In summary, the well-established filled duration illusion appeared clearly for some listeners, but did not appear for others. When the subjective durations were measured with the method of magnitude estimation, the illusion was more likely to occur; it occurred for a considerable number of participants who did not exhibit it in the method of adjustment task. The experimental method can influence the occurrence of the filled duration illusion, and the absence of the illusion in some cases and the occurrence in others may be related to the location of the listener’s attention.

Acknowledgement. This research was supported by grants to YN and KU from the Japan Society for the Promotion of Science (19103003, 20653054) and to SG from the Natural Sciences and Engineering Council of Canada.

References

Craig, J. C. (1973). A constant error in the perception of brief temporal intervals. *Perception and Psychophysics*, *13*, 99-104.

Grondin, S. (2010).Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception & Psychophysics, 72*, 561-582.

Hasuo, E., Nakajima, Y., & Ueda, K. (2011). Does filled duration illusion occur for very short time intervals? *Acoustical Science and Technology*, *32*, 82-85.

Repp, B. H. & Marcus, R. J. (2010). No sustained sound illusion in rhythmic sequences. *Music Perception*, *28*, 121-133.

Wearden, J. H., Norton, R., Martin, S., & Montford-Bebb, O. (2007). Internal clock processes and the filled duration illusion. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 716-729.

Zwicker (1969/70). Subjektive und objective Dauer von Schallimpulsen und Schallpausen [Subjective and objective duration of sound impulses and sound pauses]. *Acustica*, *22*, 214-218.

DOES REQUIRING PARTICIPANTS TO PRODUCE THE FOREPERIOD DURATION ELIMINATE THE EFFECT OF FOREPERIOD LENGTH ON A SUBSEQUENT RT TASK?

Craig Leth-Steensen & Adam Chitty
Department of Psychology, Carleton University, Ottawa, K1G 5B6, Canada
craig_leth_steensen@carleton.ca

Abstract

It is commonly found that shorter foreperiod durations yield faster RTs than longer foreperiod durations. This finding has been attributed to decreased time uncertainty for shorter foreperiods. In the present study, time uncertainty effects were eliminated by having the participants produce the foreperiod durations. Twenty psychology students participated in an experiment involving an even-odd digit RT task. Two foreperiod durations of either 2- or 8-sec in length were used, which the participants were asked to estimate themselves. Immediately following the production of these durations, the digit task was presented. The results of this experiment showed that RT to the digit task was still significantly faster in the 2-sec temporal production blocks than in the 8-sec temporal production blocks. As this RT difference between the 2- and 8-sec 'foreperiod' durations cannot be attributed to time uncertainty, these results call into question the validity of time uncertainty as an explanation for the presence of fixed foreperiod effects more generally.

The nature of the factors that affect people's ability to prepare themselves for action has always been a subject of great interest. One extremely robust effect is that increasing the length of time between the occurrence of a neutral warning signal and the occurrence of a reaction stimulus, known as the foreperiod duration, tends to slow down the time taken to respond to the reaction stimulus, but only when such durations are held constant within blocks of trials (Neimi & Naatanen, 1981). The most commonly accepted explanation for this effect is that longer foreperiod durations are associated with more time uncertainty regarding when the reaction stimulus is likely to appear. If this view is valid, then eliminating time uncertainty should eliminate the effect of foreperiod length on the time taken to respond to the subsequent reaction stimulus. This notion is tested here.

One key factor affecting a participant's response time (RT) is the degree to which they are mentally prepared at the exact moment the reaction stimulus is presented. Shorter RTs are associated with a higher degree of mental preparedness while longer RTs are associated with a lower degree of mental preparedness. In any RT experiment, there are five main elements involved (Niemi & Naatanen, 1981), specifically, the warning signal, foreperiod, reaction stimulus, response, and the inter-trial interval.

The warning signal is a stimulus preceding the reaction stimulus that signifies to the participant that the reaction stimulus is likely to occur within an allocated amount of time. This signal is typically a visual or an auditory stimulus that can vary according to certain characteristics (e.g., brightness or loudness). The foreperiod represents the time interval between the warning signal and the reaction stimulus. It is a common variable to study as this is the phase where preparatory processes take place in anticipation to react. When studying foreperiod effects, a major assumption is that if the participants can estimate the length of the foreperiod, then they can also strategically optimize their readiness to respond at the exact moment the reaction stimulus occurs. As such, better accuracy in estimating the length of the duration of the foreperiod by the participants should result in shorter RT (Niemi & Naatanen, 1981).

When studying RT and foreperiod effects, it is important to note that RT results depend first and foremost on the manner in which the foreperiod length is manipulated. That is, the foreperiod length can either be held constant in a block of trials, known as fixed foreperiods, or it can vary within a block of trials, known as variable foreperiods. It has been found that when pure blocks of fixed foreperiods are used, RT increases as the length of the foreperiod duration increases (Niemi & Naatanen, 1981). This finding has been widely accepted when dealing with fixed foreperiods (Leth-Steensen, 2009).

It is often assumed that the primary reason why shorter fixed foreperiods yield faster RTs than longer fixed foreperiods is because shorter foreperiods allow for a more accurate prediction of the timing of the impending reaction stimulus (Correa et al., 2006). In other words, under fixed foreperiod conditions when the foreperiod duration is known, longer foreperiods are associated with increased time uncertainty regarding the moment of occurrence of the reaction stimulus, which then limits the participant's level of preparedness at the exact moment the reaction stimulus is presented. This assumption is consistent with the scalar timing view of the temporal estimation process (Gibbon, Church, & Meck, 1994) and the notion that as the length of time being estimated increases, estimates of it will tend to be more variable (Klemmer, 1957).

One finding that has been found to be inconsistent with the strategic time estimation view of the fixed foreperiod effect has been RT distributional results obtained by Leth-Steensen (2009) using a digit magnitude classification task. The main finding reported in that study was that lengthening the foreperiod duration from 2- to 8-sec between blocks of trials resulted in a shift of the front ends of the RT distributions upwards. As argued by Leth-Steensen (2009) such a finding is inconsistent with the strategic time estimation account of fixed foreperiod duration effects which would predict an increase in RT variability in both the tails and the front end of the RT distribution but not a consistent RT shift.

The Current Study

In the current study, two foreperiod durations of either 2- or 8-sec in length were used across blocks of trials. However, time uncertainty for the end of the foreperiod duration length was eliminated by having the participants estimate the length of the foreperiod duration themselves. That is, at the start of each trial, the participants had to first produce a time interval either 2- or 8-sec in length. As soon as they indicated that the duration has finished, a digit stimulus appeared to which a speeded response was made. Because the digit was timed to appear exactly at the moment that the participants believed the (2- or 8-sec) duration to be up, they could never be uncertain about the moment of its occurrence. Hence, they should have been equally prepared to respond to it regardless of the length of time that had just been estimated.

Method

Participants. Twenty students from Carleton University participated for course credit. Eight participants were male and 12 female with a mean age of 25.45 years (*SD* = 10.92). Each participant performed the task individually for a 90 min. session.

Materials and Apparatus. The tasks were programmed and ran on a 486 PC using Micro Experimental Laboratory (MEL: Micro Experimental Lab System 2.0 Psychology Software Tools, Pittsburgh, PA) which regulated event sequencing, generated the stimuli, and collected the responses. Vocal responses into a microphone that was connected to a PST serial MEL response box were used to end the production of the time intervals. The stimuli for the RT task were the digits 1, 2, 3, 4, 6, 7, 8 and 9 which were presented in the centre of the computer screen in white against a black background (4 mm in width and 6 mm in height). Responses to this task were made manually by pressing one of the first two keys on the MEL response box keys using the index and middle fingers of the right hand. Participants sat about 40 cm from the screen with the vocal microphone in front of them (but not obstructing their view of the screen in any way).

Procedure. Participants first performed two blocks of 80 time production trials. The time intervals to be produced in each of these blocks were 2- or 8-sec in length (with the order counterbalanced across participants). Each trial began with a “Ready?” prompt on the screen. When participants were ready to begin producing the time interval, they were instructed to press the space bar (with their left index finger). When they felt that the length of the duration had passed, they were to say the word “Time” into the microphone. Feedback was provided if the produced duration had either underestimated or overestimated the actual duration by a certain amount. This feedback involved presenting either the phrase “Incorrect – your estimation was too short!” or “Incorrect – your estimation was too long!” for 1 sec. For the 2-sec intervals, any productions within ± 300 ms were regarded as correct and no feedback was provided (i.e., the “Ready?” signal for the next trial appeared right away). Similarly, for the 8-sec intervals, any productions within ± 800 ms were regarded as correct. These two windows were chosen because productions of the longer 8-sec interval were naturally assumed to be more variable (i.e., less sensitive) and also because a bit of pilot work indicated that these two windows provided approximately equal production “accuracy” (about 70-75%).

In the remaining four blocks, the RT task was also performed. Participants were instructed that as soon as they had finished producing each time interval (i.e., right after saying “Time”), a digit would immediately appear on the screen. They were then to classify this digit as being either even or odd by pressing the left or right response keys, respectively. Participants were asked to respond to the digits as quickly as they could without unduly sacrificing accuracy, and it was also requested of them that they keep their fingers rested on these two response keys at all times. After the digit response, feedback regarding any under- or overestimation of the time interval (only) was provided (in exactly the same manner as described above). For participants who had started the experiment by producing 2-sec time intervals, the time intervals produced in these last four blocks were 2, 8, 8, and 2 sec, respectively, For participants who had started the experiment by producing 8-sec time intervals, the time intervals produced in these last four blocks were 8, 2, 2, and 8 sec, respectively. Each of these four blocks contained 64 trials (8 per digit) although the first two were each preceded by 16 practice trials.

Results

In this experiment, only the data from the last four blocks, excluding the practice trials, were used in the following analyses. The raw data of interest were the sets of 2- and 8-sec production times and the corresponding RTs to respond to the digit stimuli in each of the 2- and 8-sec temporal production blocks.

Production Times. Before any analyses were performed on the production time data, they were adjusted for any outliers by removing any produced durations greater than 3 standard deviations above the mean of each participant's production times in each of the 2- and 8-sec temporal production conditions. As well, all production times below 400 ms were regarded as erroneous productions and were also removed. These criteria resulted in the removal of 2.8% of the production responses. Means and standard deviations of the remaining production times in each of the 2- and 8-sec temporal production conditions were then obtained for each individual.

In accordance with the mean accuracy property of scalar timing theory, the overall mean of the 2-sec temporal productions was 1988 ms whereas the overall mean of the 8-sec temporal productions was 7868 ms, both of which are very close to the actual durations (i.e., relative accuracies were -0.6% and -1.65%, respectively). In accordance with the variance property of scalar timing theory, the mean of the standard deviations for the 2-sec temporal productions was 336 ms whereas the mean of the standard deviations for the 8-sec temporal productions increased to 980 ms (although the coefficients of variation were 5.92 and 8.03, respectively, indicating that the increase in the standard deviation of the production times was not strictly proportional to the increase in the mean).

Reaction Times. Before any analyses were performed on the RT data, they were adjusted for both outliers and incorrect responses. Removing any RTs associated with incorrect answers in the digit task resulted in the deletion of 5.27% of these data. A further adjustment of the correct RT data by eliminating RT outliers greater than 3 standard deviations beyond the mean of each participant's correct RTs within each of the 2- and 8-sec temporal production conditions, resulted in the further removal of 1.84% of the correct RTs. Median correct RTs in each cell of the following design were then obtained for each individual (where one outlying cell median was replaced by the cell mean + participant mean – grand mean; i.e., which is equivalent to the grand mean + cell effect + participant effect).

A within-subjects ANOVA was performed on the digit task RT medians that involved two factors: temporal production condition (i.e., 2- or 8-sec) and digit stimulus (1, 2, 3, 4, 6, 7, 8 or 9). A significant main effect of the length of the produced duration on the digit task RTs was indeed present, $F(1, 19) = 5.308$, $p < .033$, partial $\eta^2 = .218$, where $M =$ 607 ms for the 2-sec production condition but $M =$ 671 ms for the 8-sec condition. There was also a significant main effect of the specific digit stimuli on RT, $F(7, 133) = 6.808$, $p < .001$, partial $\eta^2 = .264$, where the exact nature of these effects is evident in Figure 1. The interaction between the production condition and the digit stimulus factors was not significant, $F(7, 133) = 0.709$, $p > .50$, partial $\eta^2 = .036$.

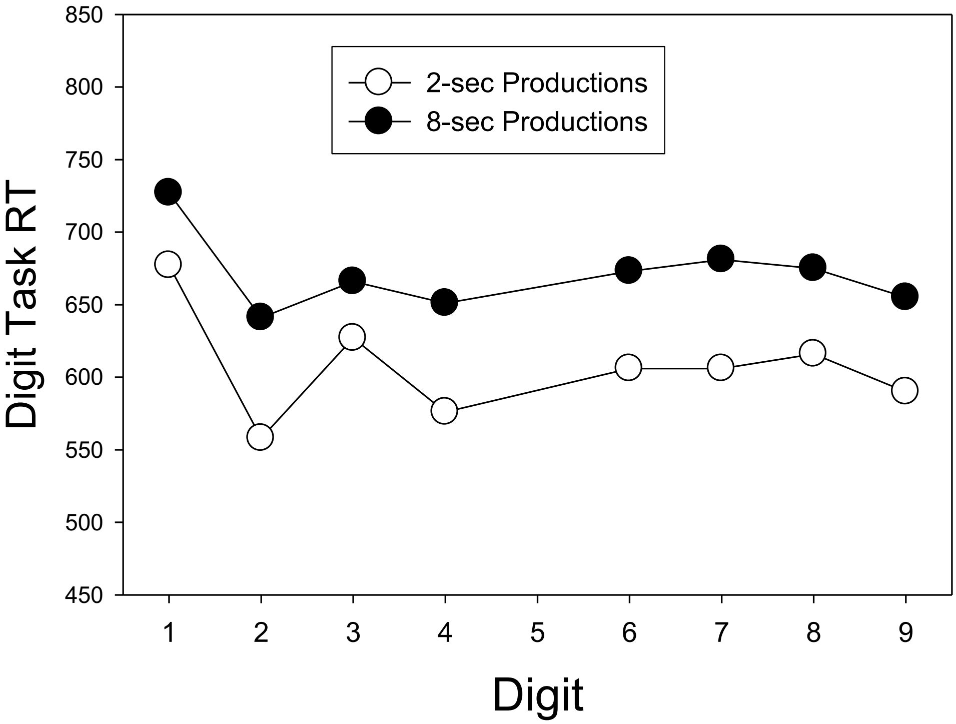


Figure 1. Digit task RT as a function of the digit and the foreperiod production condition.

Accuracies. A within-subjects ANOVA was performed on the (arc-sine transformed) accuracy of responding to the digit task that involved two factors: temporal production condition (i.e., 2- or 8-sec) and digit stimulus (1, 2, 3, 4, 6, 7, 8 or 9). A marginally significant main effect of the length of the produced duration on the digit task accuracy was present, $F(1, 19) = 3.287$, $p < .086$, partial $\eta^2 = .147$, where $M = .940$ (untransformed) for the 2-sec production condition but $M = .955$ (untransformed) for the 8-sec condition. There was a significant main effect of the specific digit stimuli on accuracy, $F(7, 133) = 9.902$, $p < .001$, partial $\eta^2 = .343$, that was analogous to that found in RT. The interaction between the production condition and the digit stimulus factors was not significant, $F(7, 133) = 0.811$, $p > .50$, partial $\eta^2 = .041$.

Discussion

The current study examined foreperiod RT effects in a completely novel fashion, namely, by having the participants estimate and produce the designated foreperiod lengths (after which time the reaction stimulus immediately appeared). This manipulation, in effect, served to eliminate the temporal uncertainty associated with the onset of the reaction stimulus (given that the foreperiod duration was always terminated by the participants themselves). Nonetheless, the RT results indicated that participants were still faster at the digit task when it followed a 2-sec temporal production “foreperiod” as opposed to an 8-sec one (i.e.. the standard foreperiod effect occurred). Importantly, this effect cannot be attributed to time uncertainty here. In turn, such a result then also calls into question the validity of this view as an explanation for the presence of fixed foreperiod effects in more standard RT studies.

In the temporal production part of this experiment, participants were asked to produce durations that were to last either 2- or 8-sec (after which they were to respond to the reaction stimulus). The results for this temporal production task indicated that participants were indeed trying to be accurate in these time productions for both the 2- and 8-sec durations given that the means of their production times were quite close to the requested time durations and the standard deviations of their production times were larger for the 8-sec productions than for the 2-sec productions (in accordance with scalar timing theory).

One final caveat regarding the RT results, though, is the fact that producing 2- and 8-sec time intervals could be assumed to consume more cognitive resources than simply waiting out such time intervals given that temporal estimation tasks have been assumed to be somewhat attentionally demanding (Block et al., 2010). However, it would also have to be assumed that attentional demands are less for the 2-sec productions in comparison to the 8-sec ones in order to account for the observed differences in digit task RTs.

To conclude, digit task RTs were faster after participant-mediated fixed foreperiod durations of 2 sec than 8 sec, which replicates the standard fixed foreperiod effect. Because time uncertainty could not possibly be the prevailing factor determining the effect obtained here, its role with respect to the more commonly observed fixed foreperiod effects might indeed be questioned.

References

Block, R. A., Hancock, P. A., & Zakay, D. (2010). How cognitive load affects duration judgments: A meta-analytic review. *Acta Psychologica*, 134, 330-343.

Correa, A., Lupianez, J., & Tudela, P. (2006). The attentional mechanisms of temporal orienting: Determinants and attributes. *Experimental Brain Research*, 169, 58-68

Gibbon, J., Church, R. M., & Meck, W. H., (1984). Scalar timing in memory. In J. Gibbon, & L. G. Allan (Eds), *Annals of the New York Academy of Sciences: Timing and time perception* , New York: New York Academy of Sciences, pp. 52-77.

Klemmer, E. T. (1957). Simple reaction time as a function of time uncertainty. *Journal of Experimental Psychology*, 54, 195-200.

Leth-Steensen, C. (2009). Lengthening fixed preparatory foreperiod durations within a digit magnitude classification task serves mainly to shift distributions of response times upwards. *Acta Psychologica*, 130, 72-80.

Niemi, P., & Naatanen, R. (1981). Foreperiod and Simple Reaction-Time. *Psychological Bulletin*, 89, 133-162.

PERFORMANCE EVALUATION OF DIVING USING THE BORG CR100 SCALE®

Elisabet Borg and Gunnar Borg
Department of Psychology, Stockholm University, SE-106 91 Stockholm, Sweden
eb@psychology.su.se

Abstract

In some sports, as for example in diving, performance is measured as a subjectively evaluated artistic gestalt. The purpose of this study was to compare the traditional scale used in competitive diving with the Borg CR100 scale®, a scale where categorical expressions are placed where they perceptually belong on a ratio scale (e.g., G. Borg and E. Borg, 2001). Two internationally recognized Swedish judges volunteered as subjects and judged a sample of 45 videotaped dives, both with the traditional scale and with the CR scale. The results show that the Borg CR100 scale® worked at least equally well as the traditional scale, even though there might have been some tendency for translation between scales.

Within most sports and athletics, performance is measured by reliable physical measures for time, length, weight or amount. For a number of sports, for example diving, figure skating, ski jumping, etc., the best performance is, however, the ideal combination of several separate parts in an artistic gestalt, and is subjectively evaluated. The scales used differ somewhat, but are usually of ordinal or "semi-interval" character.

In diving, for example, five (or sometimes seven) judges evaluate their overall impression of the technique and grace of the dive based on the starting position and the approach, the take-off, the flight, and the entry into the water. The scale used internationally up until recently had the following categories (FINA, 2002):

Completely failed	0
Unsatisfactory	0.5 – 2.0
Deficient	2.5 – 4.5
Satisfactory	5.0 – 6.0
Good	6.5 – 8.0
Very good (perfect)	8.5 – 10

The purpose of this pilot study was to compare the scale above with the psychophysical category-ratio scale developed by Gunnar and Elisabet Borg (G. Borg and E. Borg, 1994, 2001; E. Borg and G. Borg, 2002; and E. Borg, 2007), the Borg CR100 scale® (Fig. 1) in performance evaluation of diving. The CR100 scale is a general intensity scale that combines the value of Stevens' ratio scaling (Stevens, 1975) with the value of obtaining direct level estimation made possible by the categories of the scale. One of the main principles behind the construction of the CR scales, is Gunnar Borg's range model that emphasizes the need of an interindividually valid reference point, usually defined as a previously perceived maximal intensity, of, for example, perceived exertion (G. Borg, 1962, 1998).

Method

Two professional Swedish judges partook in the experiment. Both had many years of experience in judging diving internationally. The material used was 45 videotaped dives of varying difficulty from international diving contests (10 m springboard: men and women; 1 m springboard: men, from 1996 - 2003). The selection of dives was made by one of the professional judges.

The dive was presented to the judges on a TV-monitor and the video was paused directly at the finish of the dive before the result was shown. The judges then made their evaluation according to the traditional scale (TS) presented above in the introduction, where "Good" is the main reference level, and then according to the Borg CR100 scale® (Fig. 1). Maximal (100) on the CR100 scale was defined as a perfect dive. To obtain a criterion for each dive the competition result for the dive was divided by the product of the difficulty of the dive and 0.6 times the number of judges (FINA, 2002).

Results

The average results for the two judges (A and B) and the two scales, as well as for the criterion (competition results for the dives) are presented in Fig. 2. Only the upper half of each scale was used (from 4.5 on the traditional scale and from 45 on CR scale). Data distributions were somewhat more even with the CR scale, as well as average agreement between the two judges.

Individual data for both scales are shown in Fig. 3 with the criterion on the x-axis. On the traditional scale the finest increment was 0.5, whereas the Borg CR100 scale® is more finely graded. To study to what extent this was used, data from the traditional scale was multiplied by 10 and the differences between the two scales were computed. For Judge A there was a 5 point difference in 10 dives, but in 15 dives (33%) the difference was between 1 to 4 points. The corresponding values for Judge B was 3 and 4 dives respectively.

Correlations are presented in Table 1. High correlations, above 0.7, were obtained with the criterion for both judges and both scales. Inter-rater correlations were also above 0.7. The correlations between the two scales were very high, above 0.9, for both judges.

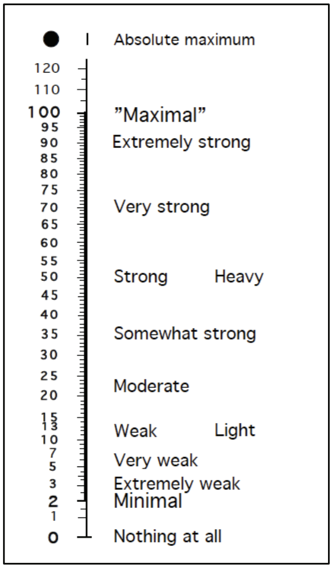


Fig. 1. The Borg CR100 scale® (© G. Borg and E. Borg, 1987, 1994, 2001, 2007).

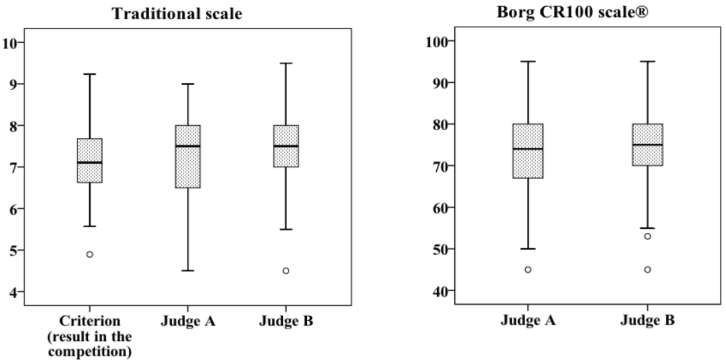


Fig. 2. Boxplot graphs (medians in the center and 25th and 75th percentiles as the edges of the box). Obtained performance evaluation of 45 dives with the traditional scale (left) and the Borg CR100 scale® (right).

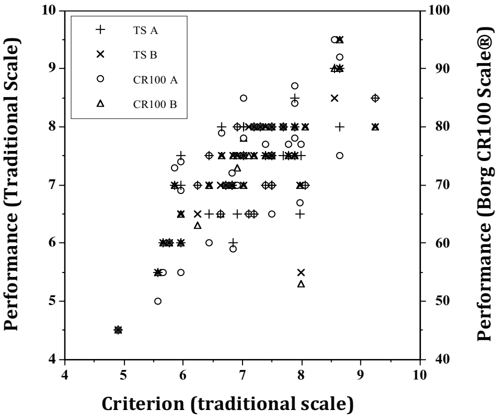


Fig. 3. Individual data for both judges and both scales.

Table 1. Correlations with the criterion and between evaluations for 45 dives ($p < .001$).

	Criterion	Judge A (TS)	Judge A (CR100)	Judge B (TS)
Judge A (TS)	.72			
Judge A (CR100)	.72	.97		
Judge B (TS)	.78	.72	.69	
Judge B (CR100)	.78	.74	.71	.99

Discussion

The purpose of this study was to investigate the possibility to apply the Borg CR100 scale® (Fig. 1) for performance evaluation in diving. As can be seen from Fig. 3 and Table 1 the Borg CR100 scale® worked equally well compared to the criterion as did the traditional scale.

The Borg CR100 scale® is more finely graded than the traditional scale. This advantage was used by Judge A in 33% of the dives, but only in a few dives by Judge B. Another advantage with the CR100 is the inclusion of a stable "ideal" dive at the maximum of the scale. The lack of this in the traditional scale may explain the small displacement between the two judges found with the traditional scale, but not with the CR100 (Fig. 2). Since this study was performed the scale used for diving, has been somewhat altered in that an additional anchor, "excellent", has been added to the top of the scale (10) (FINA, 2012).

The CR scale is a general intensity scale and the verbal anchors on the scale were not specifically chosen to suit diving. A suspicion that arises (visible also in Fig. 2) is that scale values were only translated between scales. Despite the instruction to use the verbal anchors on the Borg CR100 scale® this seems to some extent to have been the case, since the values used covered a very similar number range (separated only by a factor 10). Thus, instead of using "Strong" (50) on the CR100 scale as a correspondence to "Good" (which is probably more similar as to perceived intensity), "50" on the CR scale was used as a correspondence to 5,0 on the traditional scale ("Satisfactory"). This points to the need to further emphasize the importance of the verbal anchors on the CR100. An alternative design where only one scale was used at the time might also have reduced this tendency. The judges are, however, so used to the traditional scale, that it is very possible that this would make no difference.

There are several sports, for example diving, figure skating, ski jumping, etc., where the Borg CR100 scale® could be of value for performance evaluation. Emphasis should then be put on the importance to not just translate from one scale to the other, but of using the verbal anchors. It would also be interesting to try the Borg CR100 scale® on a larger group of judges both on recorded dives and in a live situation.

Acknowledgments

This research was sponsored by grants from the Swedish National Centre for Research in Sports. The authors want to thank the two judges, Mathz Lindberg and Jan-Erik Nilsson, as well as Anders Thomson and Emil Wiktorsson for help with data collection.

References

Borg, E. (2007). *On Perceived Exertion and its Measurement.* Doctoral dissertation. Department of Psychology, Stockholm university. http://su.diva-portal.org/smash/record.jsf?searchId=1&pid=diva2:197216

Borg, E., and Borg, G. (2002). A Comparison of AME and CR100 for Scaling Perceived Exertion. *Acta Psychologica, 109,* 157-175.

Borg, G. (1962). Physical performance and perceived exertion. *Studia Psychologica et Paedagogica,* Series altera, Investigationes XI. Lund, Sweden: Gleerup.

Borg, G. (1998). *Borg's Perceived Exertion and Pain Scales*. Champaign, IL:Human Kinetics.

Borg, G., and Borg, E. (1994). Principles and experiments in Category-Ratio scaling. *Reports from the Department of Psychology,* Stockholm university, No.789.

Borg, G., and Borg, E. (2001). A new generation of scaling methods: Level-anchored ratio scaling. *Psychologica, 28,* 15-45.

FINA PERMANTENT OFFICE (2002). *Diving manual organisation and judging.* Av. L'Avant-Poste 4, 1005 Lausanne, Switzerland.

FINA (2012) http://www.fina.org/H2O/

Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural and social prospects*. New York, Wiley.

WHAT'S NEW? NOT US

Scott Parker
American University, Washington DC
sparker@american.edu

Abstract

In twelve current History of Psychology texts, psychophysics gets no discussion for the years since about 1880. Signal Detection Theory appears in just one book (and is identified only as what APA gave John Swets an award for). Stevens' Law appears in only two books, one of those including it as a "refinement" of Fechner's Law. For comparison, five books mention two-point thresholds, five mention Weber's Law, all twelve mention Fechner's Law (often without a formula), seven mention Fechner's psychophysical "methods" (apparently methods for accomplishing something of little or no subsequent importance) and three provide Fechner's interesting biography. Have we accomplished nothing since then? Have our successes become so commonplace and familiar that they don't merit discussion? What became of us? What have we done for them lately and why don't they notice?

Many schools offer a course on the history of psychology (sometimes called History and Systems) and some require their students to take it. I inspected twelve of the recent textbooks (some that a friend who teaches the course had, and some more hurriedly at publishers' booths at conferences). In essence, as these textbooks present it, psychophysics did not contribute to the development of psychology as a science after about 1880. (I also looked at a few books from the 1990s; they look quite like the more recent ones in these respects.)

It is quite obvious by inspecting the books' topical coverage that people write books based in part on reading what the competition looks like. They all say something or other about thresholds, occasionally (Benjamin, 2007) distinguishing the Ascending and Descending Methods of Limits. Many of the 12 discuss the 2-point threshold in particular as an example of a difference threshold, before discussing Weber's Law in general. Most credit Fechner with developing "methods" for determining thresholds. Several mention Fechner's law, some mention that it's "logarithmic" and some provide a formula. So the textbooks do properly put some of psychophysics into the early history of scientific psychology. In addition, several provide some interesting biographical information on Fechner (who is, indeed, a good story) and one mentions that Fechner influenced Freud's ideas about the unconscious. None of those latter things are psychophysics, but at least they're Fechner.

But once Wundt's laboratory is established, things change. Consider the treatment of S. S. Stevens, for example. Four books mention him in connection with the rise of operationism in psychology. One mentions him in passing and only because George Miller was a research fellow in Stevens's lab. Two mention Stevens's power function and both do so where they're discussing Fechner; one of those provides a formula and the other merely says that it's a mathematical refinement of Fechner's original psychological principle. One (Chung & Hyland, 2012) says that "Fechner's work forms the basis for modern psychophysical measurement" and cites two Stevens references – "On the Psychophysical Law" and "To Honor Fechner and Repeal His Law". In no books do the discussions of Fechner's and Stevens's laws have any experimental results accompanying them – those two formulations appear to be simply theoretical exercises.

Signal Detection Theory fares even worse than does psychophysical scaling. It gets a mention in only one book, there in connection with John Swets's receiving the APA

Distinguished Scientific Contribution Award for his "extension" of SDT "as an alternative to the magnitude estimations of Fechner and Stevens". Sadly, "magnitude estimations" appear nowhere else in the book. Happily, SDT is said to treat the experimental subject as an active decision maker operating under uncertainty.

So what's our problem?

First, psychophysics suffers from having been famously characterized as boring by William James (1890), in part due to its focus on methodology. At one point (p.192) he complained that "psychology is passing into a less simple phase. Within a few years what one may call a microscopic psychology has arisen in Germany, carried on by experimental methods, asking of course every moment for introspective data, but eliminating their uncertainty by operating on a large scale and taking statistical means. This method taxes patience to the utmost, and could hardly have arisen in a country whose natives could be *bored*. Such Germans as Weber, Fechner, Vierordt, and Wundt obviously cannot". He was not alone in this diagnosis; one of psychophysics's supporters thought so too. L. L. Thurstone (1931) told the Midwestern Psychological Assocation that several years earlier "it seemed to me that psychophysics was really a very dull subject in spite of the fact that it did offer the satisfaction of clean and quantitative logic…..There is a great deal of hairsplitting about just how a limen should be determined with the greatest possible precision…And then you can find shortcuts for these methods by which you can determine somebody's limen very quickly when you are in a hurry for a limen…[But] I venture the guess that not more than perhaps half a dozen psychologists in this room have ever needed or wanted somebody's limen for anything with a high degree of precision" (pp. 249 - 250). And, indeed, the history books do reliably discuss the "methods" for determining limens. We have continued to find more efficient schemes for finding limens and discrimination functions and the like (see Kingdom & Prins, 2010, Chapter 5 for several). We are right to regard psychophysics's triumphs in this regard as important but the broad range of psychology is not so methodologically careful, nor is it interested that we are.

Second, psychophysics suffers from its primary association with the study of the senses. Sensory topics arise only little in history texts. One book discussed color vision in an early chapter (indeed, the book's first chapter not entirely about the history of philosophy) called "Physiological Influences on the Development of Psychology". That chapter discussed Galvani, Phineas Gage and Hodgkin & Huxley; the different views of Hering and of Young & Helmholtz and those of Christine Ladd-Franklin and Hurvich & Jameson. It also included the material on psychophysics, and the work of von Bekesy modernizing Helmholtz's view of pitch perception. None of those topics reappeared later. Gestalt psychology, however, does get extended coverage (sometimes entire chapters) considerably later in several books. So by and large, the history books do not take up the study of sensory and perceptual processes (though James Gibson's "ecological" approach to perception gets a few pages in a few books, mostly because it's "ecological"). The history books' inattention to sensory and perceptual topics in the 20th century surely contributes to their lack of interest in psychophysics.

If psychophysics shows up in introductory psychology textbooks it is in the introduction to the chapter(s) on sensory function where students learn about thresholds and occasionally SDT. That's where it appears in the popular introductory text by Schacter, Gilbert and Wegner (2011). That book's chapter on the senses, unfortunately, never mentions or displays any thresholds other than Galanter's (1962) list of limens specified in "familiar" physical terms.

Thurstone (1931) declared that psychophysics would be no less logical but far more interesting if its subject matter expanded from the measurement of sensory experience to matters of "more psychological significance than the sensory limen" (p.250) – in particular, attitudes. And so he proposed taking up matters of more psychological significance, e.g., asking people “Which of these two offenses do you consider in general the more serious?” rather than “Which of these two little cylinders is the heavier?” (p.250). He and Chave (1929) had already declared their system for attitude measurement to be “psychophysical” in the title of their monograph. And, indeed, there has been work along these lines on what we might call social psychophysics (see, e.g., Stevens, 1975, chapter 8; Wegener, 1982).

We in ISP often have an interest in the notion of measurement. None of the history texts mention Stevens’s classification of measurement scales nor any subsequent development of what we think of as measurement theory. Although I did not take proper notes on this topic, I know that some of the history books do discuss the psychometricians’ concepts of reliability and validity as important considerations in measurement associated with mental testing. All this is quite consistent with the treatment of measurement issues in textbooks of introductory psychology. For instance, consider the treatment of measurement in Schacter, Gilbert and Wegner (2011). The authors say that measurement requires “operational definitions” (and so they propose that we can define “happiness” by the frequency with which a person smiles or by the answer to the question, “How happy are you?”) Scale type of course goes unmentioned because in this approach there is no actual thing properly called “happiness” other than the measurement adopted in the “operational definition”. And so by metaphysical sleight-of-hand, the measurement becomes the thing it measures. Despite that, the book pulls back from that position to discuss validity as “the extent to which a measurement and a property are conceptually related”. It then discusses reliability and power (“the ability of a measure to detect the concrete conditions specified in the operational definition”). All this is accomplished in approximately three-quarters of a page. It is followed by a page-and-a-half on Demand Characteristics and another half-page on Observer Bias. What we in ISP think are the important characteristics of measurement are not major concerns in this introductory textbook nor in any others that I know of.

One important thing that unites psychophysics with many other areas of psychology is the underlying belief that subjective experience is measurable, that stimuli can be thought to have locations along some psychological dimension. As Marks (1982, p.43) put it, “The modern enterprise that Fechner may be said to have begun is the quantification of mental events.” (William James, on the other hand, thought of Fechner’s enterprise as one whose “proper psychological outcome is just nothing” [p.534]). Of course, people before Fechner thought that sort of quantification possible in one setting or another (Marks mentions Plato), but Fechner provided the first program for trying to accomplish it. And it is so intuitively appealing that it seems altogether reasonable that we can try to measure people’s impressions of occupational prestige, or seriousness of crimes, or the locations of politicians or voters along a dimension extending from “Very Conservative” to “Very Liberal”. (This discussion ignores some important differences between psychophysics and psychometrics, and it will not be clarified here.) This sort of idea has motivated a great deal of psychophysical work and much of it (at least much of the sensory and perceptual work) tells a coherent story (for a good recent summary, see Teghtsoonian, 2012).

The notion of subjective experience is what motivates much of social psychologists’ interest in such “context effects” as assimilation and contrast. Two recent social psychology books (Biernat, 2005; Stapel & Suls, 2007) take explicit note of the psychophysics ancestry of the social psychology investigations. And it is an idea for which we should receive more credit.

Psychophysics needs something of a public relations campaign. It needs to reduce the textbook emphasis on psychophysical methods for learning things and enhance attention to what we’ve learned using them. If we can do it with intrinsically interesting subject matters, so much the better. We need to speak not only of measurement accomplishments but answer the question, “Why does humanity want to know this?” Applied work using SDT or the sort of work that Howard Moskowitz (2012) does with food might help reawaken general interest in our field. So might a wider appreciation of that fact that SDT has provided informative new views of such varied phenomena as “depressive realism” (Allan, Siegel, & Hannah, 2007), and the influence on memory of “levels of processing” (Sheridan & Reingold, 2012).

This leads to what may be the largest part of our problem. Fechner thought of psychophysics as the study of the relations of body and mind, of the material and the mental. He distinguished between outer and inner psychophysics. To a considerable extent, psychophysics has been essentially subsumed into cognitive science (outer psychophysics) and behavioral neuroscience (inner psychophysics). And so even if some of our interests continue to get mentioned in the history books, our name has disappeared. For example, it is rare (I think) to see mentions of psychophysical ideas and their contributions to the ideas of the now-popular behavioral economics including the attempts to determine the “utility function”. (Poundstone [2010] is not a psychologist but his trade book on the topic makes our role in all that clear. Sadly, the psychologists and economists who think of themselves as behavioral economists seem not to.)

Fifty years ago, Gene Galanter (1962) described psychophysics as the interconnected studies of detection, recognition, discrimination, and scaling. That was a sensible way to present psychophysics as a science with a subject matter – processes that occur widely in a great variety of circumstances. Now, all that remains of that chapter is a list of catchy values for human sensory thresholds. Perhaps it’s time for a new try.

The foregoing incorporates both ideas and phrases generously supplied by Larry Marks, Bruce Schneider, Bob Teghtsoonian, and Debra Zellner.

References

Allan, L. G., Siegel, S., & Hannah, S. (2007). The sad truth about depressive realism. *Quarterly Journal of Experimental Psychology, 60*, 482 – 495.

Benjamin, L. (2006). *A brief history of modern psychology*. New York: Wiley/Blackwell.

Biernat, M. (2005). *Standards and expectancies: Contrast and assimilation in judgments of self and others.* New York: Psychology Press.

Bolles, R. C. (1993). *The story of psychology: A thematic history.* Belmont CA: Brooks/Cole.

Brennan, J. F. (1998). *History and systems of psychology 5/e*. Upper Saddle River NJ: Prentice Hall.

Chung, M. C., & Hyland, M. E. (2012). *History and philosophy of psychology*. New York: Wiley/Blackwell.

Galanter, E. (1962). Contemporary psychophysics. In *New directions in psychology* (pp.87 – 156). New York: Holt, Rinehart & Winston.

Gillham, N. W. (2001). *A life of Sir Francis Galton*. New York: Oxford University Press.

Goodwin, C. J. (2008). *A history of modern psychology* (3/e). New York: Wiley.

Greenwood, J. D. (2009). *A conceptual history of psychology.* New York: McGraw-Hill.

Hergenhahn, B. R. (2009). *An introduction to the history of psychology.* Belmont CA: Wadsworth.

James, W. (1890) *The principles of psychology, Volume 1*. New York: Holt.

Kingdom, F. A. A., & Pitts, N. (2010). *Psychophysics: A practical introduction*. London: Elsevier.

Lawon, R. B., Graham, J. E., & Baker, K. M. (2007). *A history of psychology.* Upper Saddle River NJ: Prentice-Hall.

Leahey, T. H. (1997). *A history of psychology 4/e.* Upper Saddle River NJ: Prentice-Hall.

Marks, L. E. (1982). Psychophysical measurement : Procedures, tasks and scales. In B. Wegener (Ed.), *). Social attitudes and psychophysical measurement* (pp. 43 – 68). Hillsdale NJ: Erlbaum.

Moskowitz, H. R., Beckley, J. H., & Resureccion, V. A. (2012). *Sensory and consumer research in food product design and development, 2/e*. New York: Wiley.

Pickren, W. E., & Rutherford, A. (2010). *A history of modern psychology in context*. New York: Wiley.

Poundstone, W. (2010). *Priceless*. New York: Hill and Wang.

Schacter, D. L., Gilbert, D. T., & Wegner, D. M. (2011). *Psychology 2/e* New York: Worth.

Schultz, D. P., & Schultz, S. E. (2008). *A history of modern psychology*. Belmont CA: Wadsworth.

Sheridan, H., & Reingold, E. M. (2012). Levels of processing influences both recollection and familiarity: Evidence from a modified remember‑know paradigm. *Consciousness and Cognition, 21*, 438 – 433.

Shiraev, E. (2011). *A history of psychology*. Thousand Oaks CA: Sage.

Stapel, D. A., & Suls, J. (2007). *Assimilation and contrast in social psychology*. New York: Psychology Press.

Stevens, S. S. (1975). *Psychophysics*. New York: Wiley.

Teghtsoonian, R. (2012). The standard model for perceived magnitude: A framework for (almost) everything known about it. *American Journal of Psychology, 125*, 165 – 174.

Thorne, B. M., & Henley, T. B. (2005). *Connections in the history and systems of psychology.* Boston: Houghton-Mifflin.

Thurstone, L. L. (1931). The measurement of social attitudes. *Journal of Abnormal and Social Psychology, 26*, 249-269.

Thurstone, L. L., & Chave, E. J. (1929). *The measurement of attitude: A psychophysical method and some experiments with a scale for measuring attitude toward the church*. Chicago: University of Chicago Press.

Viney, W., & King, D. B. (2003). *A history of psychology: Ideas and context* (3/e). Boston: Allyn & Bacon.

Wegener, B. (1982*). Social attitudes and psychophysical measurement*. Hillsdale NJ: Erlbaum.

NAMES WILL HURT YOU: EFFECT OF LABEL ON LIKING AND PREFERENCE

Debra A. Zellner, Kaila Hoer, and Juliann Feldman
Department of Psychology, Montclair State University, Montclair NJ 07043 USA
zellnerd@mail.montclair.edu, hoerk1@mail.montclair.edu, feldmanj4@mail.montclair.edu

Abstract

When hedonic contrast causes stimuli to become less good it also reduces subjects' preferences between the stimuli (hedonic condensation). Here we investigate whether the reduction in preference is the result of comparing the judged stimuli to the preceding context stimuli or the result of their increased negativity. Two groups smelled and rated their degree of preference between two pairs of cheeses (one group told they were smelling cheeses and the other body odor samples). They then smelled each of the four samples individually and rated the intensity and liking for the samples with the same label. There was no effect of label on intensity ratings. But subjects told that the samples were body odor liked the samples less and showed less of a preference between pairs.

When good stimuli precede hedonically neutral stimuli not only do they make those neutral stimuli less hedonically positive (negative hedonic contrast), but they also reduce the preference between them (hedonic condensation). Zellner, Allen, Henley, & Parker (2006) demonstrated this condensation effect by having subjects rate their degree of preference between paired mediocre test juices either when those juices were preceded by pairs of hedonically positive juices or when they were presented alone. Subjects who first tasted the hedonically positive juices not only reported liking the mediocre juices less than did those subjects who tasted only the mediocre juices (negative hedonic contrast), but they also reported smaller preferences between the mediocre test juices. Similar results were found with subjects evaluating the attractivenesses of pictures of birds (Zellner, Mattingly, & Parker, 2009).

Here we investigate whether the reduction in preference (hedonic condensation) that accompanies negative hedonic contrast is the direct result of comparing the judged stimuli to the preceding context stimuli or the result of their increased negativity. In order to distinguish between these two alternatives we had to have a way to reduce the hedonic value of a set of stimuli in some way other than producing negative hedonic contrast. Previous research has shown that the hedonic evaluation of odors can be reduced by labeling the odors with a label that subjects find unpleasant. For example, in a study by Herz and von Clef (2001) subjects rated a mixture of isovaleric and butyric acid as less pleasant when labeled as "vomit" than when labeled as "parmesan cheese". A similar effect was found when a mixture of isovaleric acid and cheddar cheese smell was labeled as either "cheddar cheese" or "body odor" (de Araujo, Rolls, Velazco, Margot & Cayeux, 2005).

We use this labeling technique ("cheese" vs. "body odor") to alter subjects' hedonic evaluations of a single set of stimuli by means other than hedonic contrast. Since the hedonic values of the stimuli will be shifted without the use of hedonically positive context stimuli (as would be the case in a hedonic contrast study) we will be able to determine if the hedonic condensation that accompanies negative hedonic contrast is a result of the contrast procedure or whether negative shifts in hedonic value that occur for other reasons can also produce hedonic condensation.

Method

Participants. Twenty undergraduate students from Montclair State University (4 males and 16 females) served in the experiment. Their mean age was 20.4 years. Subjects were tested individually.

Stimuli. The four stimuli were 2oz samples of four cheeses. The cheeses were Munster, Provolone, Swiss and Parmesan purchased at a local grocery store deli counter. The cheeses were wrapped in a small piece of cheesecloth secured with a small piece of masking tape. The samples were put into small glass bottles with screwcaps and allowed to reach room temperature before testing began.

Procedure. The 20 subjects were randomly assigned to one of two groups: the Cheese Group or the Body Odor Group. Subjects in the Cheese Group were told that they would be smelling and rating different cheese samples. Subjects in the Body Odor Group were told that they would be smelling and rating different samples of body odor.

The subjects were first given pairs of the cheeses to rate (Provolone paired with Swiss, Munster paired with Parmesan). They were handed two bottles at the same time and were told to smell them both and then indicate if they liked one sample more than the other. If they did they were then asked which one they preferred and how much more they liked it than the other. They indicated their degree of preference using a 10-point rating scale on which 1 indicated liking the preferred sample “slightly more”, 4 indicated “somewhat more”, 7 indicated “a lot more”, and 10 indicated “very much more” than the other sample of the pair. If a subject had no preference between a pair of odors the experimenter assigned a rating of 0 for that pair.

Subjects were then given all four samples, one at a time in random order and smelled and rated how much they liked each odor using a 201-point bipolar hedonic scale. A rating of –100 indicated that the subject thought the cheese/body odor sample was the “most unpleasant imaginable”; 0 indicated the subject found it “neither pleasant nor unpleasant”; and +100 meant that the subject found sample the “most pleasant imaginable”.

Subjects were then given the four samples a third time, one at a time in random order. They smelled and rated the intensity of each sample using a 0 (no odor) to 100 (most intense odor imaginable) scale.

Results

Preference. We calculated the average preference rating given to the two pairs of odors for each subject. Subjects in the Cheese group rated the preference for one odor over the other in the pair (M = 4.20, SD = 1.06) as significantly larger than did the Body Odor group of subjects (M = 3.05, SD = 1.09), $t(18) = 2.39$, p=.03, Cohen’s d = 1.13. See Table 1.

Liking. We calculated the average hedonic rating given to the four odors for each subject. The Body Odor subjects rated the odors (M = -30.75, SD = 28.22) as significantly less pleasant than did the Cheese subjects (M = -7.02, SD = 13.20), $t(18)$ =2.41, p=.03, Cohen’s d = 1.14. See Table 1.

Intensity. We calculated the average intensity rating given to the four odors for each subject. There was no significant difference in intensity ratings for the odors between the Body Odor subjects (*M* = 51.05, *SD* = 20.12) and the Cheese subjects (*M* = 48.38, *SD* = 8.21), *t*(18) = 0.39, *p*=.70, Cohen's *d* = 0.18. See Table 1.

Table 1. Mean preference, hedonic, and intensity ratings (and standard deviations) for the odors by the Cheese and Body Odor groups.

Group	Preference	Hedonic	Intensity
Cheese	4.20 (1.06)	- 7.02 (13.20)	48.38 (8.21)
Body Odor	3.05 (1.09)	-30.75 (28.22)	51.05 (20.12)

Discussion

This is the first demonstration that odor labels can affect the size of preference judgments. The shift in the size of the preference judgments co-occurred with a shift in the hedonic ratings. So, a reduction in hedonic rating whether it is a result of negative hedonic contrast (Zellner et al., 2006; Zellner et al., 2009), or a result of labeling, affects preference judgments in the same way, reducing them. Condensation does not just accompany negative hedonic contrast, it seems to occur whenever there is a reduction in hedonic value, whatever the cause. Stimuli that are less liked are also judged as less hedonically different.

Although both the hedonic ratings and the preference ratings of the cheese odors were reduced when the cheese odors were labeled as body odor, the intensity judgments were not affected. This suggests that the affective evaluation of the stimulus is the only thing that changes. This finding is consistent with the findings of de Araujo et al. (2005) where labeling an isovaleric acid/cheddar cheese flavor odor as "body odor" resulted in a lower hedonic rating than when the same odor was labeled as "cheddar cheese" but no effect of intensity ratings was caused by the same labeling. In addition, that study, using an event-related fMRI design, found that the rostral anterior cingulate cortex (ACC)/medial orbitofrontal cortex (OFC) was significantly more activated when the subjects thought they were rating "cheddar cheese" than when they thought they were rating "body odor". The degree of activation of these areas was correlated with hedonic ratings. These areas are different from the areas of the brain that appear to be correlated with differences in perceived intensity (piriform cortex, Rolls, Kringelbach, & de Araujo, 2003).

This suggests that the hedonic condensation effect is not the result of the same mechanism that produces hedonic contrast. One mechanism that has been posited as producing both hedonic contrast and hedonic condensation is the increase in the size of the hedonic range (Parducci, 1995). The introduction of the hedonically positive context stimuli preceding the more hedonically neutral target stimuli which occurs in studies of negative hedonic contrast increases the range of stimuli the subject experiences. It has been suggested that that increase in range might produce negative hedonic contrast and hedonic condensation (Zellner et al., 2009). In the present study there is no shift in the size of the hedonic range. The stimuli presented are always the same. Therefore, some other mechanism must explain hedonic condensation.

One possibility is that the effects on size of preference have something to do with the structure of the hedonic scale (Parker & Zellner, 1988). When hedonically good stimuli precede mediocre stimuli, the hedonic ratings of those stimuli are often pushed from slightly above to slightly below hedonic neutrality (e.g., Zellner et al., 2003). In this study, the mediocre test stimuli are moved from slightly to moderately below hedonic neutrality. It could be that people just don't put a lot of effort into discriminating between two stimuli which are hedonically negative. If the stimuli are clearly not good, people might not care which of the two is slightly better. People might not have a preference between two stimuli that they find unpleasant. The more pleasant they become, the more valuable it might be to discriminate between them and choose the best one of the pair. This possibility is roughly consistent with Wedell, Hicklin, and Smarandescu's (2007) view that attentional mechanisms govern discrimination. It requires the additional assumption that for hedonic judgments, attention is not spread evenly over the stimulus range nor governed by stimulus density but rather that particular regions of the hedonic scale are more attention-grabbing than others.

This explanation also suggests that contrast-induced shifts in hedonic evaluation of stimuli are not due simply to changes in the use of the rating scale. Instead, hedonic contrast must cause an actual shift in the perceived hedonic value of the target stimuli which results in hedonic condensation.

References

De Araujo, I.E., Rolls, E.T., Velazco, M.I., Margot, C., & Cayeux, I. (2005). Cognitive modulation of olfactory processing. *Neuron, 46*, 671-679.

Herz, R.S. & von Clef, J. (2001). The influence of verbal labeling on the perception of odors: Evidence for olfactory illusions? *Perception, 30*, 381-391.

Parker, S. & Zellner, D.A. (1988). The difference between goodness and badness. In H. Ross (Ed.), *Fechner Day 88. Proceedings of the Fourth Annual Meeting of the international Society for Psychophysics* (pp. 107-108). Stirling, Scotland: Department of Psychology, University of Stirling.

Parducci, A. (1995). *Happiness, pleasure, and judgment*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Rolls, E.T., Kringelbach, M.L., & de Araujo, I.E.T. (2003). Different representations of pleasant and unpleasant odors in the human brain. *European Journal of Neuroscience, 18*, 695-703.

Wedell, D.H., Hicklin, S.K., & Smarandescu, L.O. (2007). Contrasting models of assimilation and contrast. In D.A.Stapel & J. Suls (Eds), *Assimilation and contrast in social psychology* (pp. 45-74). New York: Psychological Press.

Zellner, D. A., Allen, D., Henley, M., & Parker, S. (2006). Hedonic contrast and condensation: Good stimuli make mediocre stimuli less good and less different. *Psychonomic Bulletin & Review, 13*, 235 – 239.

Zellner, D.A., Mattingly, M.C., & Parker, S. (2009). Categorization reduces the effect of context on hedonic preference. *Attention, Perception, & Psychophysics, 71*, 1228-1232.

Zellner, D. A., Rohm, E. A., Bassetti, T. L., & Parker, S. (2003). Compared to what? Effects of categorization on hedonic contrast. *Psychonomic Bulletin & Review, 10*, 468 – 473.

A HALF-CENTURY'S PERSPECTIVE ON BÉKÉSY TRACKING AND LOUDNESS GROWTH AT THRESHOLD

Elizabeth Rooney [1] and Mary Florentine [1,2]

Department of Speech-Language Pathology and Audiology [1]
Department of Electrical and Computer Engineering [2]
Northeastern University (106-A FR), 360 Huntington Avenue, Boston, MA 02115 USA

Rooney.e@husky.neu.edu and *m.florentine@neu.edu*

Abstract

In the more than half a century since Georg von Békésy was awarded the Nobel Prize, many of his findings have stood the test of time. Implications of the width of the excursions in the Békésy tracking procedure, however, are not as simple as described in his classic book, Experiments in Hearing (von Békésy, 1960). The interpretations of his findings have created some misunderstandings in hearing science and audiology. This article reviews a large body of data on Békésy tracking and more recent threshold, discrimination, and loudness data in order to answer the following question: Is the width of the excursions a reliable indicator of cochlear pathology?

In 1947, von Békésy described a new audiometer and claimed that this new audiometer permitted the testing of recruitment. Recruitment has been defined as a rapid growth in loudness above the threshold of hearing (Fowler,1936; Steinberg and Gardner, 1937; Brunt, 1994) and was believed to be associated with all cochlear hearing losses (Dix, Hallpike, and Hood, 1948). The audiometer was automated. As the frequency of the signal increased slowly, the subject pressed a button when the signal was audible and released it when inaudible. This pressing and releasing of the button caused the level of the signal to decrease or increase, respectively, and a visible tracing of the excursions (or trackings) between the just-audible and just-inaudible pulsed tones could be observed on recording paper that was attached to a rotating drum. This method was widely used for determining absolute threshold.

In addition to providing a measure of absolute threshold, von Békésy asserted that the width of the tracking excursions was a reliable indicator of cochlear pathology (von Békésy, 1947). His reasoning was based on the observations that cochlear hearing loss is characterized by elevated absolute thresholds and normal (or near-normal) loudness for high-level sounds. It was widely assumed that cochlear hearing loss resulted in rapid loudness growth and, accordingly, better-than-normal level discrimination. Von Békésy further assumed that the rapid loudness growth would be consistent with narrow tracking excursions and small discrimination thresholds at and near absolute threshold. Although the assumption that the width of the tracking excursions is a reliable indicator of cochlear pathology seemed reasonable at the time, there is a need to re-examine this theoretical framework with a half-century's perspective and new data.

Early experiments (1950-1975) in search of empirical support for a theoretical framework

The introduction of the Békésy audiometer sparked a flurry of debate about the clinical interpretation of the width of the tracking excursions. Von Békésy believed the width of the excursions represented the difference limen at threshold. In fact, von Békésy (1947) claimed that his audiometric procedure was "the equivalent of the recruitment test of Fowler but had the further advantage of permitting the investigation of a single ear." This claim was hotly debated. Despite a lack of empirical evidence showing that the difference limen was an accurate measure of recruitment, this idea persisted.

Several articles stated support of the claim that Békésy tracking is a test of recruitment. For example, Reger (1952) reported several ways in which the Békésy tracking procedure can be used, stating that recruitment testing was one of the main advantages to this testing system. Furthermore, he stated that "any child or adult capable of taking the test unknowingly tests himself for recruitment over the test frequency range", despite a lack of published evidence. Reger and Kos (1952) examined the relationship between temporary threshold shifts and Békésy audiograms. They report that the excursions should be examined closely and demonstrate the recruitment phenomenon. In these articles, assumptions were made about recruitment without the data to support it.

There was also discussion about variables affecting the tracking procedure itself. Lüsher (1951) reported that excursions could be affected by reaction time, concentration of the patient, and fatigue. Denes and Naunton (1950) reported that there were many factors that could affect the width of the excursions, including attention, the patient's sense of rhythm, and fatigue. [Supporting data indicating judgmental biases and criterion shifts due to the perception of a rhythmic pattern were subsequently obtained by Miśkiewicz, Buus, and Florentine (1994).] For normal listeners, the excursion widths were generally accepted to be within the 5-15 dB range (Palva, 1956; Jerger, 1960). Siegentahler (1975) examined 63 listeners with normal hearing and suggested that variability in the width of the excursion might be due to attenuation rate of the Békésy audiometer and the test frequency. In particular, the results of the study showed that using slower attenuation rates and high-frequency stimuli created smaller excursion widths and more reliable results as examined through test-retest statistical analysis using the same 63 normally hearing subjects. It is possible that any of these factors could obscure real differences between normal listeners and listeners with cochlear hearing losses. In addition to the aforementioned variables, there were reports that high-level cognitive functioning could impact the results. Harbert and Young (1968) studied 30 subjects with cochlear hearing loss and reported that the width of the excursions was influenced by the patient's intelligence, motivation, central habituation, attention, senility, and reaction time.

Some researchers believed that the width of the excursions gave an indication of variability and adaptation. Hirsh, Palva, and Goodman (1954) reported that von Békésy was actually measuring the variability about the absolute threshold and not the difference limen. Tsuiki (1966) suggested that the narrowing of the excursions was due to rapid adaptation and not recruitment. Later, Harbert and Young (1964; 1968) examined several Békésy tracings from ears with different pathologies and agreed that abnormal adaptation is the cause of the narrowing of the excursions. Whatever was being measured, by the end of the 1970s it was clear that the width of tracings, as described by von Békésy (1947), was not a good predictor of general cochlear pathology. It was clear that some of the assumptions needed to be reexamined by addressing the following questions.

Is level discrimination better for listeners with cochlear pathology than normal listeners?

The difference limen (DL) for level (aka, intensity DL or JND) is the just-noticeable difference between two audible stimuli and, accordingly, is measured at supra-threshold levels. It has been measured using different experimental paradigms (*e.g.,* two-tone comparison and amplitude modulation). One problem with using an amplitude-modulation paradigm is that the data are strongly frequency dependent for normal listeners. [For review, see Buus *et al.* (1982a,b) and

Turner *et al*. (1989).] The two-tone comparison procedure yields little frequency dependence (Jesteadt *et al*., 1977), except above 4 kHz (Florentine 1983; Florentine *et al*., 1987). Several studies (Hirsh *et al*., 1954; Jerger and Jerger, 1967; Florentine *et al*., 1979; 1993; Turner *et al*., 1989) measured the level DLs using two-tone comparison and concluded that there was little to distinguish between recruiting, non-recruiting, and normal listeners. Only Denes and Naunton (1950) reported that the DL for level increases or remains constant for recruiting ears. In the most comprehensive study of level DLs as a function of level in listeners with sensorineural hearing loss, Florentine, Reed, Rabinowitz, Braida, Durlach, and Buus (1993) showed that different audiometric configurations result in different DL functions for level. Although it was not possible to be certain of the etiology of hearing loss, some etiologies have been associated with specific audiometric configurations of hearing loss. Their results also indicate that there are large individual differences among listeners with sensorineural hearing losses and that the level DLs are not smaller than normal, except perhaps in the case of rising audiograms. The available data, therefore, are overwhelmingly inconsistent with a correlation between recruitment and smaller-than-normal DLs for level in cochlear pathology.

Is loudness at threshold the same for cochlear pathology and normal hearing?

One common assumption was that loudness at threshold was zero and, accordingly, it was displayed as zero in published figures. This assumption was never tested because it was commonly assumed that loudness below 4 dB SL could not be measured due to limitations of the available psychophysical procedures. Advancements in an innovative experimental paradigm by Buus, Müsch, and Florentine (1998) led to a new understanding of loudness at threshold. [For review, see Florentine (2009) and Florentine, Popper, and Fay, 2011).] Their measurements showed a low, but positive value of loudness at threshold. This finding is now widely accepted and it has been incorporated into the new ANSI S3.4-2007 loudness standard.

The collapse of the assumption that loudness at threshold was zero opened other assumptions to scrutiny. If loudness at threshold is truly greater than zero, it may have a value that differs among individuals. It is possible that damage to the cochlea causing hearing loss may also affect loudness perception, especially at the threshold of hearing. Buus and Florentine (2002) actually measured loudness at threshold and found that some listeners with cochlear hearing losses have abnormally elevated loudness at threshold, which is called softness imperception. Although Moore (2004) reported results that appear inconsistent with the concept of softness imperception, he only tested four listeners with one procedure. Because equal-loudness measurements are fraught with problems at low SLs (Hellman and Zwislocki, 1961), it is desirable to obtain loudness measurements using a variety of converging methods. For this reason, Florentine, Buus, and Rosenberg (2005) measured reaction times (RTs) for 200-ms tones at different frequencies. Because equal simple RTs for tones have been shown to correspond closely to equal loudness and because measurements of RTs are possible even when tones are set at threshold (for review, see Wagner *et al*., 2004; Epstein, 2011), RTs were used to indirectly assess loudness at threshold. Results for six listeners with cochlear hearing losses consistently showed faster RTs to tones at and near threshold for a frequency with elevated threshold than for a frequency with normal or near-normal threshold. This evidence supports the idea that some listeners with cochlear hearing losses have greater-than-normal loudness at threshold. Florentine, Mumby, and Cleveland (2004) further tested this concept with 22 listeners with cochlear hearing losses and the results suggest the existence of softness imperception. The majority of the data indicate that loudness at threshold has a positive value and that loudness can be greater than normal in some listeners with hearing losses of primarily cochlear origin, and that differences in test frequency do not account for this effect (Epstein and Florentine, 2006).

Is the rate of loudness growth faster in cochlear hearing losses than normal hearing?

Hellman and Zwislocki (1961) provided compelling data that loudness near threshold had an average slope of about unity or slightly larger for listeners with normal hearing. According to a summary figure from Buus and colleagues (1998) published decades later, Hellman and Zwislocki's finding was supported by a number of studies that measured the loudness of low-level tones using a variety of methods.

Although it had been assumed that cochlear pathology leads to an abnormally rapid growth of loudness immediately above an elevated threshold, this assumption is not supported by empirical data. The rate of loudness growth with increasing level at and near threshold for listeners with sensorineural hearing losses is not radically different from listeners with normal hearing (Hellman and Zwislocki, 1964; Hellman, 1997, 1999; Buus and Florentine, 2002; Moore, 2004; Smeds and Leijon, 2011).

Current View

The previous sections of this article show that none of the three assumptions have stood up to empirical testing. Level discrimination is not generally better for listeners with cochlear pathology than listeners with normal hearing, and can often be worse. Loudness at threshold is highly likely to be different for cochlear pathology than for listeners with normal hearing. The rate of loudness growth with increasing level near threshold is usually not much greater for cochlear hearing loss than for normal hearing. It should be noted that these are general comments based on group data, and that there are large individual differences in auditory processing among listeners with hearing losses that have been considered to be of predominantly cochlear origin. It is becoming increasing clear that not all listeners with cochlear hearing losses have the same abnormal loudness growth. When data for listeners with hearing losses are averaged, important individual differences are averaged out, as pointed out by Marozeau and Florentine (2007). In fact, Marozeau and Florentine (2007) examined data from past studies to test two loudness-growth theories: softness imperception and recruitment. Five different studies using different methods to obtain individual loudness functions were used: absolute magnitude estimation, cross-modality matching with string length, categorical loudness scaling, loudness functions derived from binaural loudness summation, and loudness functions derived from spectral summation of loudness. Results showed large individual differences. Some of the individuals demonstrated softness imperception, others showed recruitment, and there were some who did not fit into either group yet showed abnormal loudness growth. This research indicates that there is large individual difference and one method of determining loudness growth or one loudness growth theory does not account for all hearing losses of primarily cochlear origin.

Conclusion

Von Békésy's (1947) audiometer was a very innovative apparatus for measuring auditory thresholds. There is no evidence, however, of its ability to determine the presence of recruitment and cochlear pathology. The claim that his audiometer was "…the equivalent of the recruitment test of Fowler but has the further advantage of permitting the investigation of a single ear" has not stood the test of empirical inquiry. Although von Békésy was incorrect in his assumptions about recruitment and determining loudness growth at threshold, his theoretical work provided a basis for empirical research, which has led us to our current understanding of loudness growth functions.

Acknowledgments

NIH-NIDCD provided funding that made part of the work possible. The authors are grateful to Michael Epstein and Mindy Rosenberg for helpful comments.

References

Békésy, G.V. (1947). A new audiometer. *Acta Otolaryngol.*45: 411-422.

Békésy, G.V. (1960). *Experiments in Hearing*. New York: McGraw-Hill.

Brunt, M.A. (1994). Tests of cochlear function. In: Katz, J. (ed.) *Handbook of Clinical Audiology, 4th ed.* Williams and Wilkins, Baltimore, MD. p. 165-175.

Buus, S., and Florentine, M. (2002). Growth of loudness in listeners with cochlear hearing losses: Recruitment reconsidered. *J. Assoc. Res. Otolaryngol.*, 3, 120-139.

Buus, S., Florentine, M., & Redden, R. (1982a). The SISI test: A review, Part I. *Audiology*. 21: 273-293.

Buus, S., Florentine, M., & Redden, R. (1982b). The SISI test: A review, Part II. *Audiology.* 21: 365-385.

Buus, S., Müsch, H., & Florentine, M. (1998). On loudness at threshold. *J. Acoust. Soc. Am.*, 104, 399-410.

Denes, P., & Naunton, R.F. (1950). The clinical detection of auditory recruitment. *J. Laryngol. Otol*. 65: 375-398.

Dix, M.R., Hallpike, C.S., & Hood, J.D. (1948). Observations upon the loudness recruitment phenomenon, with especial reference to the differential diagnosis of disorders of the internal ear and VIII nerve. *Proc. R. Soc. Med.* 41: 516-526.

Epstein, M. (2011). *Correlates of Loudness.* In: Florentine, M., Popper, A.N., & Fay, R.R. (ed.), *Loudness.* New York: Springer.

Epstein, M. & Florentine, M. (2006). Reaction time to 1- and 4-kHz tones as a function of sensation level in listeners with normal hearing. *Ear & Hearing.* 27: 424-429.

Florentine, M. (1983). Intensity discrimination as a function of level and frequency and its relation to high-frequency hearing. *J. Acoust. Soc. Am.* 74: 1375-1379.

Florentine, M. (2009). Advancements in psychophysics lead to a new understanding of loudness in normal hearing and hearing loss. In: Elliott, M.A., Antonijević, S., Berthaud, S., Mulcahy, P., Martyn, C., Bargery, B., & Schmidt, H. (ed.) *Fechner Day 2009*, Proceedings of the 25 Annual Meeting of the International Society for Psychophysics: Snap Printing, Galway, pp.83-88.

Florentine, M., Buus, S., & Mason, C.R. (1987). Level discrimination as a function of level for tones 0.25 to 16 kHz. *J. Acoust. Soc. Am.* 81: 1528-1541.

Florentine, M., Buus, S., & Rosenberg, M. (2005). Reaction-time data support the existence of softness imperception in cochlear hearing loss. In McAdams, S., Cheveigne, A., Collet, L., and Pressnitzer, D. (ed.), *Auditory Signal Processing: Physiology, Psychoacoustics and Models*. New York: Springer.

Florentine, M., Mumby, S., & Cleveland, S. (2004). FYI: Softness imperceptions. *Hearing Health*, 20, 31-34.

Florentine, M., Popper, A.N. & Fay, R.R. (ed.) (2011). *Loudness.* New York: Springer.

Florentine, M., Reed, C., Durlach, N.I., & Braida, L.D. (1979). Intensity discrimination and loudness matches in subjects with sensorineural hearing loss. in *Speech Communication Papers Presented at the 97th Meeting of the Acoustical Society of America.* Wolf, J.J. & Klatt, D.H. (ed.) (Acoustical Society of America, New York), pp. 575-578.

Florentine, M., Reed, C.M., Rabinowitz, W.M., Braida, L.D., Durlach, N.I., & Buus, S. (1993).

Intensity perception. XIV. Intensity discrimination in listeners with sensorineural hearing loss. *J. Acoust. Soc. Am.* 94: 2575-2586.

Fowler, E.P. (1936). A method for the early detection of otosclerosis: A study of sounds well above threshold. *Archives Otolaryngol. Head Neck Surg.* 24: 731-741.

Harbert, F. & Young, I.M. (1964). Audiologic findings in Ménière's Syndrome. *Acta Oto-lanyngologica.* 57: 145-154.

Harbert, F. & Young, I.M. (1968). Clinical application of Békésy audiometry. *Laryngoscope.* 78: 487-497.

Hellman, R. P. (1997). Growth of loudness in sensorineural impairment: Experimental results and modeling implications. In Jesteadt, W. (ed.) *Modeling sensorineural hearing loss* (pp. 199-212). Mahwah, NJ US: Lawrence Erlbaum.

Hellman, R. (1999). Cross-modality matching: A tool for measuring loudness in sensorineural impairment. *Ear & Hearing*. 20: 193-213.

Hellman, R.P. & Zwislocki, J.J. (1961). Some factors affecting the estimation of loudness. *J. Acoustic. Soc. Am.* 33: 687-694.

Hellman, R. P., & Zwislocki, J. J. (1964). Loudness Function of a 1000-cps Tone in the Presence of a Masking Noise. *J. Acoust. Soc. Am.* 36: 1618.

Hirsh, I.J., Palva, T., & Goodman, A. (1954). Difference limen and recruitment. *A.M.A. Archives Otolaryngol.* 60: 525-540.

Jerger, J. (1960). Békésy audiometry in analysis of auditory disorders. *J. Speech Hear. Res.* 3: 275-287.

Jerger, J. & Jerger, S. (1967). Psychoacoustic comparison of cochlear and VIIIth nerve disorders. *J. Speech Hear. Res.* 10: 659-688.

Jesteadt, W., Weir, C.C., & Green, D.M. (1977). Intensity discrimination as a function of frequency and sensation level. *J. Acoust. Soc. Am.* 61: 169-171.

Lüsher, E. (1951). The difference limen of intensity variations of pure tones and its diagnostic significance. *J. Laryngol. Otol*. 65: 486-510.

Marozeau, J., and Florentine, M. (2007). Loudness growth in individual listeners with hearing losses: A review. *J. Acoust. Soc. Am*. 122: EL81-EL87.

Miśkiewicz, A., Buus, S., & Floretine, M. (1994). Auditory facilitation: Procedural or sensory effect. *J. Acoust. Soc. Am.* 93: 1429-1434.

Moore, B. (2004). Testing the concept of softness imperception: loudness near threshold for hearing-impaired ears. *J. Acoust. Soc. Am.* 115: 3103-3111.

Palva, T. (1956). Recruitment tests at low sensation levels. *Laryngoscope.* 66: 1519-1540.

Reger, S.N. (1952). A clinical and research version of the Békésy audiometer. *Laryngoscope*. 62: 1333-1351.

Reger, S.N. & Kos, C.M. (1952). Clinical measurements and implications of recruitment. *Annals Otol. Rhinol. Laryngol.* 61: 810-823.

Siegnethaler, B.M. (1975). Reliability of Békésy excursion size among normal adults. *J. Auditory Res.* 5: 111-114.

Smeds, K. & Leijon, A. (2011). *Loudness and Hearing Loss.* In: Florentine, M., Popper, A.N., & Fay, R.R. (ed.), *Loudness.* New York: Springer.

Steinberg, J.C. & Gardener, M.B. (1937). The dependence of hearing impairment on sound intensity. *J. Acoust. Soc. Am.* 9: 11-23.

Tsuiki, T. (1966). Role of auditory adaptation in Békésy audiometry. *International J. Audiology.* 5: 139-141.

Turner, C.W., Zwislocki, J.J., & Fillion, P.R. (1989). Intensity discrimination determined with two paradigms in normal and hearing-impaired subjects. *J. Acoust. Soc. Am.* 86: 109-115.

Wagner, E., Florentine, M., Buus, S., & McCormack, J. (2004). Spectral loudness summation and simple reaction time. *J. Acoust. Soc. Am.* 116: 1681-1686.

A BRIEF TRIP INTO THE HISTORY OF PSYCHOPHYSICAL MEASUREMENT

Sergio Cesare Masin
University of Padua, Italy
scm@unipd.it

Abstract

Psychophysical measurement was first used for scientific purposes more than 2000 years ago. Its development is overviewed to the present days with emphasis on the most promising lead.

Psychophysical measurement was used for scientific purposes for the first time by Hipparchus of Rhodes around 150 BCE. We reach this conclusion as follows.

In his encyclopedia *Naturalis Historia* (about 80 CE) Pliny the Elder wrote the following: "Hipparchus the foresaid Philosopher (a man never sufficiently praised, as who proved the affinitie of starres with men, and none more than he, affirming also, that our soules were parcell of heaven) found out and observed another new starre engendred in his time, and by the motion thereof on what day it first shone, he grew presently into a doubt, Whether it happened not very often that new starres should arise? and whether those starres also mooved not, which we imagine to be fast fixed? The same man went so farre, that he attempted (a thing even hard for God to performe) to deliver unto posteritie the just number of starres. Hee brought the said starres within the compasse of rule and art, devising certaine instruments to take their severall places, and set out their magnitudes: that thereby it might be easily discerned, not only whether the old died, and new were borne, but also whether they moved, and which way they tooke their course? likewise, whether they encreased or decreased? Thus he left the inheritance of heaven unto all men, if any one haply could be found able to enter upon it as lawfull heire" (Book II, Ch. 26, in Holland, 1601).

Thus Hipparchus recorded positions and magnitudes of stars to allow posterity to determine not only whether old stars had died and new stars were born but also whether stars had moved and whether their magnitude had changed. Pliny the Elder says that Hipparchus measured star position by instruments. He does not say how Hipparchus measured stellar magnitude. We can obtain the missing information from Ptolemy's star catalogue published in the *Almagest* (about 150 CE) since this catalogue used most probably earlier work by Hipparchus that is now lost. The catalogue gives coordinates and magnitudes of many stars. Stellar magnitude is "the class" (Book 7, Ch. 4, in Toomer, 1984) to which a star belongs in terms of perceived brightness. As we know, stellar magnitudes vary from I (brightest) to VI (dimmest).

Hipparchus could only measure perceived brightness since he had no photometer. For this he resolved to use category rating, apparently the most natural method of mental measurement. Variants of this method were subsequently used to measure magnitudes of stars in the telescope (Hearnshaw, 1996).

In 1740, Celsius and Tulenius were the first to obtain photometric measures of relative starlight intensity (Weaver, 1946) but it was only in the early 1800's that John Herschel (1829, p. 182) and Steinheil (1837) could provide sufficiently accurate measures. Their measurements showed for the first time that the relation between rated stellar brightness and physical relative starlight intensity was approximately logarithmic (Hearnshaw, 1996, p. 76).

In 1840, Plateau measured perceived lightness using his well-known method of bisection (Plateau, 1872). The method consists in defining an initial sensory interval delimited by two largely different perceived magnitudes of a sensation and by having a person produce equidifferent perceived magnitudes that divide the initial interval in equal subintervals. Each of these subintervals is taken as the mental unit of measurement.

Fechner (1860) proposed the following method for measuring those perceived sensations that co-vary with a known physical variable. For each fixed value of the physical variable, S, one determines the increment Δs of S that produces the smallest possible difference in the sensation. The smallest possible difference in sensation is assumed to be invariant with S. One determines the best-fitting function relating Δs to S, called the Weber function. With S_0 denoting the minimum S that evokes a sensation, one adds Δs to S_0, to the S resulting from this first addition, to the S resulting from this second addition, and so on, each time using the Weber function to select the Δs to be added to a new S. The number of additions of Δs to S_0 necessary to reach the S that produces the perceived magnitude being measured is the measure of this magnitude. Each addition defines one mental unit of measurement.

In 1887, Fechner argued that the rating method, the bisection method, and his own method produce acceptable measures of perceived sensation. He also argued that his own method should be preferable since it produces ratio-scale measures while the rating and bisection methods produce interval-scale measures (Scheerer, 1987).

Merkel (1888) and Fullerton and Cattell (1892) proposed the method of measurement in which a person selects a variable stimulus such that its perceived magnitude is in a fixed ratio with that of a standard stimulus. The perceived magnitude of the standard defines the mental unit of measurement. The bisection and Merkel's methods assume people's ability to judge the equalities of differences and of ratios of perceived magnitudes, respectively.

In 1921, Brown and Thomson set forth the central idea of the method of measurement today called nonmetric scaling: "To take a simple example, suppose five quantities a, b, c, d, e have really the measures 10, 16, 20, 31, 32." Have a person ignorant of these measures rank first differences $|a - b|$, $|b - c|$, $|c - d|$, ... , second differences $|d - c| - |b - a|$, $|c - b| - |e - d|$, $|b - a| - |c - b|$ ... , or even third differences. "If now we could have all these [rankings] we could space out the original quantities very closely indeed to their true positions. This can be best seen by attempting to alter some one of the values while leaving all [rankings] unaltered. Make d, for example, 29 instead of 31 and although the order a, b, c, d, e is unchanged, and also the order of the first differences, that of the second differences is completely altered (Brown & Thomson, 1921, p. 12)." The method is prohibitive since it requires a large number of stimulus trials even using only first differences (Shepard, 1966).

The methods of bisection, of Fechner, and of Merkel can only apply to mental variables that co-vary with a known physical variable. The usefulness of these methods is thus very limited. On the other hand, the rating method applies to any mental variable.

In 1929, these facts must have prompted Richardson to propose direct numerical magnitude estimation (Richardson, 1929a; Richardson & Ross, 1930) and graphic rating (Richardson, 1929b) to measure any mental magnitude. He measured strength of imagery by magnitude estimation, and saturation of red by graphic rating. Before, De Marchi (1925) used magnitude estimation to measure visual dot density. Magnitude estimation assumes that people can judge sensory ratios. Since the 1930s, Richardson's methods and variants thereof have been widely used up to today (Gescheider, 1997; Marks & Algom, 1998; Stevens, 1975).

The validity of the above methods depends on the truth at least of the most relevant assumptions on which they are based. The problem is that this truth is hard to ascertain. Most relevant assumptions are the abilities to equalize perceived differences in rating, bisection, Fechner's method, and nonmetric scaling and to judge sensory ratios in Merkel's and Richardson's methods and variants thereof. It is believed that one can test these assumptions by first axiomatically formalizing the operations that underlie the methods and then use these formalizations to draw empirically testable logico-mathematical consequences (Luce, 1972, 2002). Examples are the axiomatic formalizations of bisection (Pfanzagl, 1959), Fechner's method (Falmagne, 1985), and ratio judgment (Narens, 1996) among many others.

Unfortunately, tests derived from axiomatic formalizations involve serious difficulties. (*i*) Doubts about the truth of assumptions are transferred to the logico-mathematical consequences of the formalization. That is, the conclusion is reached that a method is valid or invalid based only on the trust one is willing to put in the correctness of the formalization. For example, Pfanzagl (1959) gave an axiomatic formalization of the bisection operation yielding the logico-mathematical consequence called bisymmetry condition. For about 50 years it has been given for granted that empirically testing the bisymmetry condition was fundamental to establish the validity of the method (Falmagne, 1974; Luce & Galanter, 1963). Instead, this condition is totally irrelevant for the purpose of testing this validity: it applies indifferently to any relative magnitude a person arbitrarily chooses to divide an interval (Masin & Toffalini, 2009). (*ii*) Tests of axioms derived from axiomatic formalizations are ordinal tests. They suffer from order effect. For example, given the sensory intensities a, b, and c, Fagot and Stewart (1969) had persons judge the ratios $R_{ab} = a / b$, $R_{bc} = b / c$, and $R_{ac} = a / c$. Consistent ratio judgments imply $R_{ac} = R_{ab} \cdot R_{bc}$. It turned out that $R_{ac} \neq R_{ab} \cdot R_{bc}$. This inequality could depend on order effects rather than revealing inability to judge sensory ratios. Control of order effects is inherently flawed since we ignore how the size of effects varies with stimulus intensity and with presentation order. (*iii*) Various other arguments conclude that the significance of axiomatic formalizations in psychology is virtually nil (Anderson, 1981, pp. 349–353; Cliff, 1992; Estes 1975; Schönemann, 1994)

How can one then determine the validity of a method of psychophysical measurement given that all available evidence indicates that tests based on axioms are insufficient? One answer comes from functional measurement theory (Anderson, 1982, pp. 246–251).

Our scientific knowledge about nature accrues through a continuous process of formulating tentative theories and critically testing them by selected methods of measurement. Although theories may survive critical tests, they remain theories and can only asymptotically be established as true by this converging corroboration (Popper, 1963). Since a law is a theory about a mathematical relation between variables and is tested by measurement methods that one selects among various other possible methods, validating the law and selecting its method of measurement are two related aspects of the same process of corroboration (Ellis, 1966). A method of measurement is tentatively valid when it yields measures for the variables involved in a (tentatively valid) law that are in the same mathematical relation as that which defines the law. The method is increasingly corroborated as it progressively agrees with other newly discovered laws. The following is an example of this corroboration process, analogous to the one in Anderson & Cuneo (1978). For other validation tests see Anderson (1996, pp. 94–96).

A large body of data from many judgment tasks indicates that people integrate information using mental operations such as adding, multiplying, averaging, weighted averaging, etc. (Anderson, 1981, 1996). It is theorized that since we are evolutionarily adapted to the everyday empirical world we integrate information about empirical physical phenomena using mental operations for information integration that match the mathematical relations between the variables involved in the respective empirical physical laws, as if we intuitively knew these laws (Anderson, 1983). This theory and its measurement methods are simultaneously validated by a corroboration process whose initial steps are as those exemplified next.

Consider a flat object on a horizontal board with object and board covered with sandpaper. The minimum force necessary to slide the object on the board is proportional to the sum of the grit numbers of the sandpapers of object and board. For each factorial combination of these grit numbers, and with object and board kept separate, Corneli and Vicovaro (2007) had persons rate the imagined friction of the object on the board after each person had felt with their fingers how coarse the surfaces of object and board were. Figure 1a shows mean

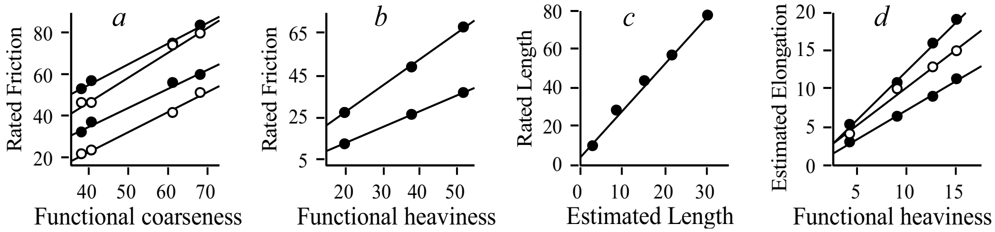


Fig. 1. Mean rated imagined friction against mean object functional coarseness (*a*) and functional heaviness (*b*), parameter: board coarseness (Corneli & Vicovaro, 2007); (*c*) mean rated length against mean estimated length (Masin, 2008); (*d*) mean estimated imagined elongation against mean object functional heaviness, parameter: spring length (Cocco & Masin, 2010).

rated friction against mean object functional coarseness for each board coarseness. [The mean of the mean ratings corresponding to each object coarseness is a mean functional measure of that felt object coarseness (Anderson, 1982, p. 58)]. The results agree with physical law.

These results tentatively support the theory that people implicitly know the additive physical law of friction and simultaneously validate the rating method used to measure imagined friction. The results suggest that ratings are linear measures of imagined friction.

These conclusions are tentative since they rest on one step only of the converging corroboration process. Some have misinterpreted that this process ends with this first step. For example, Gigerenzer and Richter (1990) argued that the same results may obtain if people multiply felt object coarseness by felt board coarseness and if ratings are logarithmic rather than linear measures of imagined friction. The next step overcomes this misinterpretation.

The minimum physical force needed to slide the object on the board equals the product of object weight by the friction coefficient. For each factorial combination of object weight and board coarseness, and with object and board kept separate, Corneli and Vicovaro (2007) had persons rate the imagined friction of the object on the board after they hefted the object and felt how coarse the board surface was. Figure 1b shows mean rated friction plotted against mean object functional heaviness for each board. The results agree with physical law.

These results further support the validity of the aforesaid theory and its measurement method—both tests involved the same method and the same measured variable but a different cognitive law. They reconfirm that ratings are linear measures: had ratings been logarithmic measures, factorial curves would have been parallel rather than being divergent.

Methods that yield linear measures are equivalent. This equivalence may hold only for some tasks. For example, ratings of average lightness are linear and nonequivalent to magnitude estimates (Weiss, 1972). On the other hand, for lengths in the range 2–68 cm, the results in Figure 1c show that ratings and magnitude estimates of apparent length are equivalent measures (Masin, 2008). Length estimation can thus be used to validate the above theory.

For a spring of length L hanging vertically, a load of weight W suspended from its lower end causes spring length to increase from L to $L + E$. The elongation E is proportional to the product $L \cdot W$. For different factorial combinations of L and W, Cocco and Masin (2010) had persons lift a load with their hands and, simultaneously, look at a spring and rate the imagined elongation of the spring that would occur in case the load was suspended on the lower end of the spring. Figure 1d shows the factorial curves relating mean estimated elongation to mean load functional heaviness for different Ls. The results agree with physical law.

These results reconfirm the validity of the aforesaid theory of implicit physical knowledge and its measurement method. They also further reconfirm that ratings and magnitude estimates of length are linear measures. This process of corroboration continues.

Convergent corroboration appears to be the only viable process of validation of psychophysical measurement. It is consequently desirable that more attempts are made at discovering and interrelating new cognitive laws such as those described above.

References

Anderson, N. H. (1981). *Foundation of information integration theory*. New York: Academic Press.

Anderson, N. H. (1982). *Methods of information integration theory*. New York: Academic Press.

Anderson, N. H. (1983). Cognitive algebra in intuitive physics. In H.-G. Geissler, H. F. J. M. Buffart, E. L. J. Leeuwenberg, & V. Sarris (Eds.), *Modern issues in perception* (pp. 229–253). Amsterdam: North-Holland.

Anderson, N. H. (1996). *A functional theory of cognition*. Hillsdale, NJ: L. Erlbaum Associates.

Anderson, N. H., & Cuneo, D. O. (1978). The height + width rule in children's judgments of quantity. *Journal of Experimental Psychology: General*, *107*, 335–378.

Brown, W., & Thomson, G. H. (1921). *The essentials of mental measurement*. London: Cambridge University Press.

Cliff, N. (1992). Abstract measurement theory and the revolution that never happened. *Psychological Science*, *3*, 186–190.

Cocco, A., & Masin, S. C. (2010). The law of elasticity. *Psicológica*, *31*, 647–657.

Corneli, E., & Vicovaro, M. (2007). Intuitive cognitive algebra of sliding friction. *Teorie & Modelli*, *12*, 133–142.

De Marchi, S. (1925). La valutazione di collettività. In [No Author] *Atti del IV Congresso Nazionale di Psicologia* (pp. 131–134). Firenze: Bandettini.

Ellis, B. (1966). *Basic concepts of measurement*. Cambridge: Cambridge University Press.

Estes, W. K. (1975). Some targets for mathematical psychology. *Journal of Mathematical Psychology*, *12*, 263–282.

Fagot, R. F., & Stewart, M. (1969). Tests of product and additive scaling axioms. *Perception & Psychophysics*, *5*, 117–123.

Falmagne, J. C. (1974). Foundations of Fechnerian psychophysics. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical* psychology (Vol. 2, pp. 127–159). San Francisco, CA: Freeman.

Falmagne, J. C. (1985). *Elements of psychophysical theory*. New York: Oxford University Press.

Fechner, G. T. (1860). *Elemente der Psychophysik*. Leipzig: Breitkopf und Härtel.

Fullerton, G. S., & Cattell, J. M. (1892). *On the perception of small differences with special reference to the extent, force and time of movement*. Philadelphia: University of Pennsylvania Press.

Gescheider, G. A. (1997). *Psychophysics: The fundamentals*. Mahwah, NJ: Erlbaum.

Gigerenzer, G., & Richter, H. R. (1990). Context effects and their interaction with development: Area judgments. *Cognitive Development*, *5*, 235–264.

Hearnshaw, J. B. (1996). *The measurement of starlight: Two centuries of astronomical photometry*. Cambridge: Cambridge University Press.

Herschel, J. F. W. (1829). Observations with a 20-feet reflecting telescope.—Third series; containing a catalogue of 384 new double and multiple stars; completing a first thousand of those objects detected in a course of sweeps with that instrument: together with observations of some previously known. *Memoirs of the Astronomical Society of London*, *3*, 177–213.

Holland, P. [Translator] (1601). *The historie of the world.* London: Adam Islip.

Luce, R. D. (1972). What sort of measurement is psychophysical measurement? *American Psychologist*, *27*, 96–106.

Luce, R. D. (2002). A psychophysical theory of intensity proportions, joint presentations, and matches. *Psychological Review*, *109*, 520–532.

Luce, R. D., & Galanter, E. (1963). Psychophysical scaling. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology* (Vol. 1, pp. 245–307). New York: Wiley.

Marks, L. E., & Algom, D. (1998). Psychophysical scaling. In M. H. Birnbaum (Ed.), *Measurement, Judgment, and Decision Making* (pp. 81–178). New York: Academic Press.

Masin, S. C. (2008). Is self-estimated linear length linear? In B. A. Schneider, B. M. Ben-David, S. Parker, & W. Wong (Eds.), *Proceedings of the twenty-fourth annual meeting of the International Society for Psychophysics* (pp. 265–270). Toronto, Canada: International Society for Psychophysics.

Masin, S. C., & Toffalini, E. (2009). On the bisection operation*. Attention, Perception, & Psychophysics*, *71*, 1566–1663.

Merkel, J. (1888). Die Abhängigkeit zwischen Reiz und Empfindung. *Philosophische Studien*, *4*, 541–594.

Narens, L. (1996). A theory of magnitude estimation. *Journal of Mathematical Psychology*, *40*, 109–129.

Pfanzagl, J. (1959). A general theory of measurement applications to utility. *Naval Research Logistics Quarterly*, *6*, 283–294.

Plateau, J. (1872). Sur la mesure des sensations physiques, et sur la loi qui lie l'intensité de ces sensations à l'intensité de la cause excitante. *Bulletin de l'Académie royale des Sciences, des Lettres et des Beaux-Arts de Belgique*, *2e série*, *33*, 376–388.

Popper, K. R. (1963). *Conjectures and refutations: The growth of scientific knowledge*. London: Routledge and Kegan Paul.

Richardson, L. F. (1929a). Imagery, conation, and cerebral conductance. *Journal of General Psychology*, *2*, 324–352.

Richardson, L. F. (1929b). Quantitative mental estimates of light and colour. *British Journal of Psychology*, *20*, 27–37.

Richardson, L. F., & Ross, J. S. (1930). Loudness and telephone current. *Journal of General Psychology*, *3*, 288–306.

Scheerer, E. [Translator] (1987). My own viewpoint on mental measurement (1887): Gustav Theodor Fechner. *Psychological Research*, *49*, 213–219.

Schönemann, P. H. (1994). Measurement: The reasonable ineffectiveness of mathematics in the social sciences. In I. Borg & P. Ph. Mohler (Eds.), *Trends and perspectives in empirical social research* (pp. 149–160). Berlin: de Gruyter.

Shepard, R. N. (1966). Metric structures in ordinal data. *Journal of Mathematical Psychology*, *3*, 287–315.

Steinheil [C. A.] (1837). Elemente der Helligkeits-Messungen am Sternenhimmel. *Abhandlungen der mathematisch-physikalischen Classe der Königlich Bayerischen Akademie der Wissenschaften*, *2*, 1–140.

Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects.* New York: Wiley.

Toomer, G. J. [Translator] (1984). *Ptolemy's Almagest*. London: Duckworth.

Weaver, H. F. (1946). The development of astronomical photometry. *Popular Astronomy*, *54*, 211–230.

Weiss, D. J. (1972). Averaging: An empirical validity criterion for magnitude estimation. *Perception & Psychophysics*, *12*, 385–388.

CROSS-MODALITY IN COMPARISONS OF SUCCESSIVE STIMULI

Åke Hellström
Caroline Cederström
Department of Psychology, Stockholm University, SE-106 91 Stockholm, Sweden
hellst@psychology.su.se

Two experiments were performed to study effects of modality, temporal position, and their interaction on comparisons of successive stimuli. In Experiment 1, intramodal (tone-tone and line-line) and crossmodal (tone-line and line-tone) stimulus pairs, with two interstimulus intervals (ISIs), were presented. Participants indicated which stimulus was the "stronger." Time-order effects (TOEs) were studied using the classic D% measure as well as weighting coefficients from Hellström's sensation weighting model. TOEs were found in both intramodal and crossmodal comparisons. The classic pause-function (more negative TOE with longer ISIs) was found in all pair types except tone-line. In Experiment 2, participants indicated which of two lines was the longer, or which of two tones was the louder. Intramodal anchors, crossmodal anchors, or no anchors were interpolated between the stimuli. Anchoring tended to yield negative TOEs and to diminish the first stimulus' weight. Intramodal anchoring of tone-tone pairs yielded low stimulus weights, suggesting stimulus interference.

It might seem that comparing two stimuli is possible only in terms of an obviously shared attribute. We can compare a high-pitched and a low-pitched tone for their loudness, or a red and a black line for their length. But can we compare a line and a tone for their magnitude?

Stimulus comparison is not free from complications even with homomodal stimuli, as has been known since the days of Fechner (1860). In the comparison of two successive or simultaneous stimuli, systematic space- and time-order errors (SOEs, TOEs) occur, in that even with identical stimuli observers tend to perceive one as being of greater magnitude than the other. It has been found (Hellström, 1979; Michels & Helson, 1954) that a major clue to understanding TOEs and SOEs is given by considering the perceived stimulus difference not in terms of subtraction, but as generated by a mechanism analougous to linear regression with different weights for the two stimuli. This type of model was elaborated into the *sensation-weighting model* (Hellström, 1979, 1985, 2003), which in its simplified form can be written

$$d_{12} = W_1\,\psi_1 - W_2\,\psi_2 + U, \qquad (1)$$

where ψ_1 and ψ_2 are the sensation magnitudes of the compared stimuli, W_1 and W_2 the stimulus weights, and U the intercept.

Crossmodal matching of magnitudes of, for instance, loudness and brightness, has long been practiced (Marks, 1974; Marks, Szczesiul, & Ohlott, 1986) yielding results that can largely be predicted using Stevens' (1957) power law for the matched continua. It has been suggested (Spence, 2011) that stimuli in different sensory modalities can be matched in terms of some amodal feature. It should therefore be of interest to study what happens in the comparison of heteromodal stimuli in the context of a TOE experiment (Exp. 1).

Earlier research (Ellis, 1933; Pratt, 1936) studied the effects of anchor stimuli interpolated between the compared stimuli. The results were interpreted in terms of assimilation of the 1st stimulus to the anchor (Pratt) or contrast of the 2nd stimulus to the anchor (Ellis). Pratt showed that an interpolated low-level anchor yielded a more negative TOE (1st stimulus underestimated relative to 2nd) than a zero-level anchor (i.e., no anchor at all). Oppenheimer et al. (2008) found that anchors in one modality influenced estimates in another modality: Long visual anchors increased numerical estimates of temperature, whereas short anchors decreased them. It is natural to ask what effect crossmodal anchors might have on the TOE. This was here investigated (Exp. 2) using the classic *D%* measure (% ’1st greater’ - % ’2nd greater’) as well as SW coefficients W_1 and W_2, for a detailed description of TOEs.

Experiment 1

Method. Participants were 11 men and 38 women aged 18-45 years ($M = 27.5$, $SD = 6.4$). All reported normal hearing and vision. ***Apparatus*****.** The experiment was run in a sound-attenuated room and controlled by a PsyScope 2.2.5 PPC script (Cohen et al., 1993). Stimuli were presented using a microcomputer with a 27” display unit and calibrated loudspeakers, which were placed on either side of the display unit. Viewing distance was ca 45 cm.

Stimuli. There were 25 tone-tone pairs (1000-Hz), combining five mean sound pressure levels (SPLs), 76.3, 77.1, 77.9, 78.7, and 78.5 dBA with five levels of within-pair SPL difference, ±3.2, ±1.6, and 0 dBA (the same pairs as were used by Hellström, 1979). Similarly, there were 25 line-line pairs, with five mean lengths, 110, 120, 130, 140, and 150 mm, combined with five levels of within-pair difference, ±40, ±20, and 0 mm. There were also 25 tone-line pairs and 25 line-tone pairs, obtained by combining the tone and line stimuli using a common level measure from -4 to +4 (see Table 1).

Procedure. Instructions were displayed, followed by eight familiarization trials, two for each modality combination. Then “Trial:“ was displayed for 725 ms, followed by a 250-ms interval until the stimulus pair was presented. Each stimulus had a duration of 250 ms. The pair was followed by “Which stimulus was the stronger?” – to be taken as generic stimulus strength. Three adjacent keys were used, labelled “first,” “second,“ and “same.” Each block was made up of a set of 25 stimulus pairs, presented twice; the set was first presented in pseudorandom order with a 400-ms ISI and then in another order with a 2000-ms ISI. The intertrial interval was, with equal probability, 1000 ms or 2000 ms. Before each block the participants were notified of the modal combination. There were four block orders; the two intramodal blocks always preceded the two crossmodal blocks, for familiarization with the stimulus ranges. The four block orders were distributed evenly across participants.

*Results***.** Crossmodal TOEs are not ’errors,’ as there is no objectively ’true’ match between a line length and a tone loudness, but they may show up as effects of ISI or modal order on *D%* as well as *W* values. The subjective stimulus difference (d_{12}), was estimated by scaling the responses: *first* = +100, *second* = -100, *equal* = 0. The mean d_{12} value for all the stimulus pairs was interpreted as the classic percent difference (*D%*) measure. For each participant and condition, the data were submitted to the SPSS procedure Regression, independent variables being the 1st and the 2nd stimulus level (-4 to +4), and the dependent variable being d_{12}. The standardized (β) coefficients were interpreted as *W* values in the simplified SW model (Equation 1). The results are presented in Table 1.

Percent difference (D%). The *D%* values were submitted to a 2 x 2 x 2 repeated measures ANOVA. The factors were modality of 1st stimulus, modality of 2nd stimulus, and ISI (400 or 2000 ms). There was an effect of the 1st modality, $F(1,48) = 5.473$, $p = .024$, $\eta^2_p = .102$: $D\%_{\text{line}} < D\%_{\text{tone}}$. There was also an effect of the ISI, $F(1,48) = 50.368$, $p < .001$, $\eta^2_p = .512$: $D\%_{2000\text{ ms}} < D\%_{400\text{ ms}}$ (the exception being tone-line, reflected by interactions).

W values. The *Ws* were submitted to 2 x 2 x 2 repeated measures ANOVAs. The factors were modality (tone-tone or line-line for intramodal; tone-line or line-tone for crossmodal), temporal position (1st or 2nd), and ISI (400 ms or 2000 ms). For the *intramodal pairs* there was an effect of modality, $F(1,48) = 30.583$, $p < .001$, $\eta^2_p = .389$: $W_{\text{line-line}} > W_{\text{tone-tone}}$, and also effects of ISI, $F(1,48) = 10.119$, $p = .003$, $\eta^2_p = .174$, $W_{2000\text{ ms}} > W_{400\text{ ms}}$, and of position, $F(1,48) = 5.094$, $p = .029$, $\eta^2_p = .096$, $W_1 < W_2$. Modality x Position x ISI interacted, $F(1,48) =$

Table 1. *Experiment 1. Stimulus Pairs, Indicating Mean Level and Level Difference*

1st stimulus Level	dBA	mm									
4	81.1	170					2/4				
3	80.3	160				1/4		2/2			
2	79.5	150			0/4		1/2		2/0		
1	78.7	140		-1/4		0/2		1/0		2/-2	
0	77.9	130	-2/4		-1/2		0/0		1/-2		2/-4
-1	77.1	120		-2/2		-1/0		0/-2		1/-4	
-2	76.3	110			-2/0		-1/-2		0/-4		
-3	75.5	100				-2/-2		-1/-4			
-4	74.7	90					-2/4				
		dBA	74.7	75.5	76.3	77.1	77.9	78.7	79.5	80.3	81.1
		mm	90	100	110	120	130	140	150	160	170
		Level	-4	-3	-2	-1	0	1	2	3	4
							2nd stimulus				

10.424, $p = .002$, $\eta^2_p = .178$. For the *crossmodal pairs* Position had a significant effect, $W_1 < W_2$, $F(1,48) = 24.258$, $p < .001$, $\eta^2_p = .336$. Position x ISI interacted, $F(1,48) = 18.854$, $p < .001$, $\eta^2_p = .282$, as well as Modality order x Position, $F(1,48) = 6.915$, $p = .011$, $\eta^2_p = .126$.

Discussion. In the SW model, a W value is interpreted as reflecting the amount of information obtained from the stimulus in the comparison (Hellström, 1985; Patching, Englund, & Hellström, in press). Line-line yielded higher Ws than tone-tone. This may be due to the lines being easier to discriminate. A general effect of temporal position was found ($W_1 < W_2$) – here implying that TOEs exist also in crossmodal comparisons. For *D%*, except for tone-line, there was an effect of ISI: the longer ISI yielded more negative *D%*s (the classic *p*-function). The results indicate that TOEs do occur in crossmodal comparisons, and suggest that the process of comparing two stimulus magnitudes is to some extent amodal, that is, involves general magnitude processing, independent of sense modality.

Experiment 2

Method. Participants were 15 men and 33 women ($M_{age} = 28$, $SD_{age} = 8.2$). All reported normal hearing and vision. The *apparatus* was the same as in Exp. 1. The *task* was to judge which of two lines was the longer, or which of two tones was the louder. The *design* had six blocks: two blocks without anchors and four blocks with an interpolated anchor between the stimuli; in two blocks this was of the same modality as the compared stimuli and in two blocks it was of a different modality. The anchor had two levels: high and low. *Stimuli.* The basic design was the same as in Exp. 1 (Fig. 1). There were 25 line-line pairs, combining five levels of line-length difference (±20, ±10, and 0 mm; in terms of level, ±4, ±2, 0) with five mean line lengths (95, 100, 105, 110, and 115 mm; in terms of level, ±2, ±1, 0). The lines were black on a white background. There were also 25 tone-tone (1000 Hz) pairs, with five SPL differences (±3.2, ±1.6, and 0 dBA) in combination with five mean SPLs (68.4, 69.2, 70.0, 70.8, and 71.6 dBA). There were also 25 tone-line pairs and 25 line-tone pairs, obtained by combining the line and tone stimuli to the same level differences and mean levels as for the

Table 2. *Experiment 1. Mean % Correct, D%, Mean W_1, Mean W_2 (*SD*s in Parentheses)*

1st stim.	ISI ms	2nd stim.	% Correct	% "Equal"	*D%*	W_1	W_2
Tone	400	Tone	67.8	39.1	5.55 (77.9)	0.50 (0.20)	0.48 (0.16)
Tone	2000	Tone	71.2	36.6	-13.14 (78.6)	0.50 (0.14)	0.58 (0.18)
Tone	400	Line	70.0	29.6	-2.20 (83.9)	0.44 (0.20)	0.56 (0.14)
Tone	2000	Line	71.2	31.3	-0.41 (82.9)	0.41 (0.20)	0.63 (0.14)
Line	400	Tone	72.2	26.4	-5.06 (85.7)	0.55 (0.13)	0.47 (0.22)
Line	2000	Tone	67.3	31.0	-15.76 (81.6)	0.46 (0.21)	0.53 (0.20)
Line	400	Line	93.4	15.8	-1.22 (91.8)	0.56 (0.15)	0.61 (0.08)
Line	2000	Line	85.0	24.9	-11.43 (85.9)	0.62 (0.09)	0.62 (0.07)

line-line and tone-tone pairs. The short and long line anchors were 75 and 135 mm; the soft and loud tone anchors were 63.4 and 76.4 dBA.

Procedure. Instructions were displayed: "You are to compare line-lengths and tone-loudnesses. In one block you will compare two line-lengths presented sequentially, and you are then to state which line was the longer. In a second block you are to compare two tone-loudnesses presented sequentially, and you are then to state which was the louder. Other blocks are similar except there will be an additional stimulus (either line or tone) between the stimuli in the pair. You are to compare the first and the last stimulus and ignore the interpolated stimulus." There were 6 practice trials (one from each block). The command "Be prepared:" was displayed for 725 ms followed by a 250-ms interval until either a stimulus dyad or triad was presented, followed by "Which line was the longer?" or "Which tone was the louder?". Three adjacent response keys were labeled "first," "same," and "last." The participants were to press "same" if they could not find another answer. Each stimulus lasted 150 ms, and the interval before and after the anchor stimulus was 1000 ms. Thus the ISI between the compared stimuli was 2150 ms (also with no anchor). The intertrial interval alternated pseudorandomly between the two values, 500 and 1000 ms. Before each block the participant was notified, for example, "You will now compare two line-lengths with a tone presented in between." The blocks and the anchor levels were presented in pseudorandom orders. The two blocks with no anchor always came first.

Results. The subjective difference score, d_{12}, was estimated as in Exp. 1. The data for each participant and condition were submitted to SPSS Regression. The independent variables were the level (-4 – +4) of the 1st and the 2nd stimulus; the dependent variable was d_{12}. The standardized regression coefficients (β values) were interpreted as *W*s in the simplified SW model (Equation 3). The results are presented in Table 3.

Anchor effects on D%. Repeated-measures ANOVAs with multivariate tests were conducted to compare the TOE measure *D%* in the five anchoring conditions. For *tone-tone D%* differed across anchoring conditions, $F(4,44) = 3.550$, $p = .014$, $\eta^2_p = .244$: *D%* was lower with all anchors than with the null anchor (soft tone, $p = .046$; loud tone, $p = .009$, long line, $p = .001$, short line, $p = .016$). An ANOVA of the *anchored* tone-tone conditions yielded no significant results.

For *line-line D%* differed across conditions, $F(4,44) = 11.526$, $p < .001$, $\eta^2_p = .512$. *D%* was more negative with the short line anchor than with the null anchor, $p < .001$. For the *anchored line-line* conditions there were effects of anchor level, $F(1,47) = 12.265$, $p = .001$, $\eta^2_p = .207$: $D\%_{\text{low}}$ (-17.347) < $D\%_{\text{high}}$ (-16.042); and of anchor modality, $F(1,47) = 16.971$, $p < .001$, $\eta^2_p = .265$: $D\%_{\text{line}}$ (-18.917) < $D\%_{\text{tone}}$ (-14.472). Anchor level x Anchor

Table 3. Experiment 2. Obtained TOE Measures

1st stimulus	Anchor	2nd stimulus	% correct	% ”same”	Mean $D\%$ (SD)	Mean W_1 (SD)	Mean W_2 (SD)
Tone	None	Tone	76.8	30.8	4.41 (83.1)	0.53 (0.12)	0.61 (0.10)
Tone	Soft tone	Tone	49.1	51.9	-2.25 (69.3)	0.05 (0.23)	0.02 (0.27)
Tone	Loud tone	Tone	48.6	54.7	-5.60 (81.8)	-0.03 (0.24)	0.09 (0.25)
Tone	Short line	Tone	75.3	27.5	-3.50 (85.1)	0.49 (0.11)	0.62 (0.12)
Tone	Long line	Tone	75.1	27.3	-2.00 (85.3)	0.44 (0.14)	0.65 (0.11)
Line	None	Line	77.5	26.9	-12.08 (84.7)	0.57 (0.13)	0.57 (0.14)
Line	Soft tone	Line	70.25	35.1	-12.18 (93.3)	0.48 (0.16)	0.58 (0.14)
Line	Loud tone	Line	70.4	33.75	-9.92 (80.8)	0.46 (0.15)	0.58 (0.14)
Line	Short line	Line	68.5	31.25	-26.08 (78.7)	0.44 (0.20)	0.53 (0.14)
Line	Long line	Line	67.8	33.5	-10.83 (80.9)	0.39 (0.16)	0.58 (0.14)

modality interacted, $F\ (1,47) = 11.715, p < 0.001, \eta^2_p = .200$.

Discriminability. As an index of discriminability, the values of $M_W = (W_1 + W_2)/2$ were submitted to repeated measures ANOVAs. For *tone-tone*, M_W differed across the five anchoring conditions, $F(4,44) = 112.224$, $p < .001$, $\eta^2_p = .911$. The tone anchors yielded worse discriminability than the null anchor, ps < .001. For *line-line*, M_W likewise differed across the five conditions, $F(4,44) = 4.329$, $p = .005$, $\eta^2_p = .282$. M_W was larger for the null anchor than for the soft-tone anchor, $p = .025$, the loud-tone anchor, $p = .005$, the short-line anchor, $p = .001$, and the long-line anchor, $p < .001$.

Weighting balance. As an index of weighting balance, the values of $W_1 - W_2$ were submitted to repeated measures ANOVAs. For *tone-tone*, $W_1 - W_2$ differed across the five anchoring conditions, $F(4,44) = 7.705$, $p < .001$, $\eta^2_p = .412$. For the soft-tone anchor, $W_1 - W_2$ was greater (positive) than with the null anchor, $p = .023$, and with the long-line anchor it was more negative, $p < .001$. For *line-line*, $W_1 - W_2$ differed across the five conditions, $F(4,44) = 5.070$, $p = .002$, $\eta^2_p = .315$. For all *anchored* conditions $W_1 - W_2$ was negative (ps < .02), whereas with the null-anchor it was zero.

Discussion. For *line-line*, the low-level intramodal anchors yielded more negative $D\%$ than the null-anchor, in accordance with Pratt (1936). The crossmodal soft-tone anchor did not exert the same pull as the intramodal short-line anchor. Also, an intramodal anchor generally yielded lower W values than a crossmodal anchor. This suggests that crossmodal anchors do not interfere with the comparison process as much as intramodal anchors.

The tone-anchored *tone-tone* conditions had very low Ws, few correct responses (ca 32%), and high percentages of “same/don’t know” answers (ca 50%). There were also low $D\%$ values and much smaller TOEs than in the other conditions. Thus, a tone anchor drastically impaired tone discrimination. This suggests that the anchor blocked the memory of the 1st tone and/or distracted attention to the 2nd tone. As this effect was much weaker for the line-anchored *line-line* conditions, it suggests different mechanisms for discriminating between auditory and between visual stimuli.

For *tone-tone*, W_1-W_2 was more negative with the crossmodal (visual) anchor than with the intramodal (auditory) anchor or the null anchor. Conversely, for the line-line comparison W_1-W_2 was less negative with the crossmodal (auditory) anchor than with the intramodal (visual) anchor, but more negative than with the null anchor. In accordance with Oppenheimer et al. (2008) the results show that anchors in one modality can influence comparisons of stimuli in another modality.

General Discussion and Conclusion

Our results show that it is indeed possible to compare a tone and a line for their ”strength,” and that TOEs show up in these comparisons. Likewise, an interpolated anchor of another modality can affect comparisons of successive stimuli. Generally, an intra- as well as crossmodal anchor rendered W_1-W_2 more negative, which may be due to interference with memory of the 1st stimulus. The drastic effect in the direction of a more negative TOE of interpolating a short line in line-line comparison is in line with Pratt (1936): The decrease in W_1, rather than increase in W_2, suggests retroactive assimilation rather than proactive contrast. According to SW theory, assimilation should increase with the reduction of W_1 (Hellström, 1985). No corresponding effect was found with crossmodal anchors. However, for tone-tone comparisons, all anchors changed the TOE (*D%*) from positive to negative. This might be due to assimilation to a softer reference level, as conceived in the full version of the SW model (Hellström, 2003; Patching, Englund, & Hellström, in press).

Author Note

This paper is based on data from the 2nd author’s M.Phil. thesis, *Comparing Apples and Oranges: The Effect of Sense Modality on the Comparison Process*, Stockholm University, 2012.

References

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavior Research Methods, Instruments & Computers, 25*, 257-271.

Ellis, H. D. (1973). Proactive effects of interpolated anchors. *Journal of Experimental Psychology, 98*, 233-238.

Fechner, G. T. (1860). *Elemente der Psychophysik* [Elements of psychophysics]. Leipzig, Germany: Breitkopf & Härtel.

Hellström, Å. (1979). Time-errors and differential sensation weighting. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 460-477.

Hellström, Å. (1985). The time-order error and its relatives: mirrors of cognitive processes in comparing. *Psychological Bulletin, 97*, 35-61.

Hellström, Å. (2003). Comparison is not just subtraction: Effects of time- and space-order on subjective stimulus difference. *Perception & Psychophysics, 65*, 1161-1177.

Marks, L. E. (1974). On associations of light and sound: The mediation of brightness, pitch, and loudness. *American Journal of Psychology, 87,* 173–188.

Marks, L. E., Szczesuil, R., & Ohlott, P. (1986). On the cross-modal perception of intensity. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 517-534.

Michels, W. C, & Helson, H. (1954). A quantitative theory of time-order effects. *American Journal of Psychology, 67,* 327-334.

Oppenheimer, D. M., LeBoeuf, R., & Brewer, N. T. (2008). Anchors Aweigh: A demonstration of cross-modality anchoring and magnitude priming. *Cognition, 106,* 13-26.

Patching, G. R., Englund, M. P., & Hellström, Å. (in press). Time-and space-order effects in timed discrimination of brighness and size of paired visual stimuli. *Journal of Experimental Psychology: Human Perception and Performance.*

Pratt, C. C. (1936). The law of disuse. *Psychological Review, 43,* 83-93.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception & Psychophysics, 73*, 971–995. DOI 10.3758/s13414-010-0073-7

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64*, 153-181.

FROM MATHEMATICS TO PSYCHOPHYSICS:
DAVID HILBERT AND THE “FECHNER CASE”

Mauro Antonelli and Verena Zudini
Department of Psychology, University of Milano-Bicocca, Milan, Italy
mauro.antonelli@unimib.it, *verena.zudini1@unimib.it*

Abstract

David Hilbert (1862-1943), the great mathematician from Königsberg and Göttingen, is a relevant figure in the science of the nineteenth and twentieth centuries, who marked the development of not only mathematical but also physical activities. We present here a sketch of Hilbert's figure and work, in particular of his contribution to the debate which ensued after the publication of Elemente *and the program of psychophysics by Fechner. In his lecture course* Logische Principien des mathematischen Denkens*, held at the University of Göttingen in 1905, Hilbert proposed an axiomatization of psychophysics. This formulation may be seen as an interesting case of Hilbert's famous axiomatic approach in natural science.*

Introduction

Fechner's work immediately had great impact on the scientific community: For the first time, with *Elemente der Psychophysik* (1860), a rigorous project of empirical and experimental research, founded and guaranteed by the possibility of measuring phenomena, was started in psychology. Since its publication, *Elemente der Psychophysik* has provoked, as is understandable considering the importance of such a project, a wide and lively debate among scholars (see Brožek & Gundlach, 1988; Heidelberger, 1993; Murray, 1993; Zudini, 2009, 2011). The model proposed in *Elemente* became that of reference: a model to criticize, correct, or confute, in the methodological aspects of its (empirical and mathematical) procedures or even in its psychophysical, physiological, or, in a strict sense, psychological value itself; in certain cases, it was a model to reject in a radical way, on the basis of the assumption that it was impossible to measure sensations and, in general, psychical magnitudes and therefore to make a scientific study on them.

The debate about Fechner's work was initially developed around three types of questions: The first concerned how correct the derivation of the law proposed by Fechner was, starting from the experimental data and from the mathematical tools he used; the second related to the very nature of the law. A third type of problem dealt with the possibility of measuring sensations and psychical magnitudes generally. Many scholars, from very different disciplines, studied the “Fechner case” and the discipline of psychophysics. Among them, there were the greatest figures of the science of the nineteenth and twentieth centuries, such as Hermann von Helmholtz, Ewald Hering, Ernst Mach, Wilhelm Wundt, Franz Brentano, Joseph Antoine F. Plateau, Joseph Delboeuf, and David Hilbert.

Hilbert, a great mathematician

David Hilbert (1862-1943), born in Königsberg (now Kaliningrad), birthplace of Immanuel Kant and known in mathematics for the “bridge problem”, was professor first at the University of Königsberg itself, then, from 1895, at the University of Göttingen, where he taught until the end of his career. Hilbert was one of the greatest mathematicians at the turn of the nineteenth and twentieth centuries, whose name is now linked with the concept of "Hilbert space", crucial in functional analysis.

Hilbert was important for significant contributions in several areas of mathematics and physics, from invariant theory, to algebraic number fields, calculus of variations, integral equations, mathematical physics, logic and theory of demonstration, as well as foundations of geometry and mathematics generally. In particular, Hilbert was the founder of "meta-mathematics" (i.e. the study of the accuracy of the methods used in mathematics); he was the head of the so-called "formalist school" – the school of thought which saw mathematics as a set of formal systems. Opposed to this was the "intuitionist school", founded by the Dutch mathematician and philosopher Luitzen Egbertus Jan Brouwer (1881-1966), traditionally referred to as "L. E. J. Brouwer", according to whom language and logic are not the presuppositions of mathematics, which instead has its origin in intuition that makes its concepts and deductions immediately clear.

Hilbert's enthusiasm for mathematics and his confidence in its vitality and potential are evident in his famous address "Mathematical Problems" delivered to the Second International Congress of Mathematicians held in Paris in 1900, a great opportunity to review the issues and problems of science pending at the turn of the century. Hilbert's optimistic contribution, containing a list of 23 problems, a veritable research program for the "coming generations", i.e. for the mathematicians of the new century, has played a determining role in marking the development of not only mathematical but also physical activities. According to Hilbert, the fundamental dynamics of the development of mathematics and, at the same time, the impetus to the process of mathematization of the other sciences lay in the continuous interaction between the free creations of reason and the knowledge of the phenomena of the external world. The rigor of the demonstrations, that peculiar quality of mathematics understood by Hilbert as "a general philosophical need of our reason", proved to be necessary in dealing with both issues of analysis and those that originated in the external world, the world of empirical experience.

The axiomatic approach in mathematics

Hilbert had given a "taste" of his approach already in his studies on the foundations of geometry, the subject of some of his lecture courses and of the work *Grundlagen der Geometrie*, first published in 1899 and revised several times over the years. *Grundlagen der Geometrie* is fundamental in the evolution of geometrical thought: In it Hilbert proposes to establish for geometry a complete and as simple as possible system of axioms and to deduce from it the most important geometrical propositions, highlighting the significance of the various groups of axioms and the extent of the conclusions which are drawn from them.

Grundlagen is testimony to the new way of conceiving geometry (as a hypothetical-deductive system) which had been emerging in the second half of the nineteenth century and which resulted necessarily in a new way of looking at the definition of the objects of the geometry itself. Hilbert's procedure is different from that of Euclid: It does not give a definition of the objects which are taken into consideration or clarify their nature at the beginning of the treatment; this is done later through the statement of the axioms and not through the names given to the objects themselves or via a reminder of the experience and the external world. Such an attitude is similar to what happens when we invent a game (e.g. cards): The actual nature of the cards (or their value) is specified by the rules of the game which we want to play more than by the figures printed on them; the definition is given implicitly by the statement of the rules and not by a phrase indicating the nature "per genus et differentiam", according to the canons of classical logic. According to this view, the object called "line" by the Euclidean geometry is not the same object which is called "line" by the non-Euclidean geometry, because the axioms of the former geometry are different from those of the latter. It is the axioms that give the rules of the game which we have to use with the objects of the geometry under consideration.

In the usage of undefined concepts (such as point, line, plane, ...) and in the fact that their properties are established only by the axioms, Hilbert follows the German mathematician Moritz Pasch (1843-1930), author of *Vorlesungen über neuere Geometrie* (1882). According to this tradition, it is not necessary to assign any explicit meaning to undefined concepts. These elements (point, line, plane, ...) could then be substituted, as Hilbert said, by tables, chairs, tankards, ... The axioms are not self-evident truths, but implicit definitions of the primitive terms which they contain; they must be considered arbitrary, even if, actually, they are suggested by experience.

After giving an example of an axiomatic construction, Hilbert addresses the problems associated with such a construction (in general, for any axiomatic system), i.e. the consistency and the mutual independence of the axioms. With regard to the consistency, that is, the coherence, he proposes to refer the coherence of the geometry to the coherence of the system of real numbers. Regarding the issue of the independence of the axioms, his discussion is of considerable interest because, when an axiom is proved as independent, at once the legitimate existence of the corresponding "non-…" geometry is suggested (for example, proving the independence of the axiom of parallels is equivalent to proving the possibility of a non-Euclidean geometry). The proof procedure is substantially similar to that applied when proving the compatibility of the axioms, i.e. by using a numerical model.

With his *Grundlagen*, Hilbert played an important role as a mathematical logician. The guarantee of the logical compatibility of the axioms of a certain group is given by Hilbert by building up every time models borrowed from other fields of mathematics. The statement of the logical compatibility of the foundations and the processes of algebra and arithmetic is thus responsible for ensuring the compatibility of the axioms of geometry. This procedure of reference, however, can not be infinite; at some point we must stop and look for a "foothold". Starting from this problem, which concerned above all the foundations of arithmetic and analysis, Hilbert developed the program of creating a "*Beweistheorie*" ("proof theory"), which envisaged the construction of a system of logical procedures that could justify the classical mathematical procedures and were justifiable themselves through finite procedures and above criticism. In other words, the aim was that of constructing a "meta-mathematics", i.e. a meta-theory which had as object mathematics and its methods and was unassailable. This project of creating a "*Beweistheorie*" – which constituted a large part of Hilbert's research program on the foundations of mathematics and was considerably developed – was frustrated by the Austrian mathematician and logician Kurt Gödel (1906-1978), who gave a fatal blow to the attempts of formalist mathematics in 1931 with his famous eponymous theorem. Gödel's contribution showed that in a theory containing arithmetic we could construct a formula whose truth is impossible to be proved and which thus would be an undecidable sentence. This would indicate the incompleteness of the theory, namely its inability to rule on all statements which it could express, regardless of the number of axioms placed at the beginning of the treatment.

The axiomatic approach in natural science and in psychophysics

In Hilbert's conception, the criterion of truth and existence of mathematical objects was then the demonstration of the consistency of the axioms and the theorems derived from them.

From this perspective, every theory was nothing but a kind of frame, a scheme of concepts together with their necessary mutual relationships, applicable to infinite systems of fundamental objects. It was sufficient that the relations between them were established from the axioms for obtaining all the propositions of the theory. The axiomatic method – and this was, according to Hilbert, its essential quality – highlighted the deductive pattern, the dependency link between axioms and theorems.

The research program proposed by Hilbert involved applying the axiomatic method to all branches of physics where mathematics played a dominant role. The aim was clearly stated in Problem 6 presented at the Congress in Paris: mathematical treatment of the axioms of physics, in particular axiomatization of those parts, such as mechanics and the theory of probability, where mathematics was essential. Here was evident the connection with the theories of probabilistic nature on kinetics of gases developed by Rudolf Clausius and Ludwig Boltzmann, as well as with the research on the principles of mechanics conducted by Ernst Mach and Boltzmann himself, in the context of an increasingly marked interest in theoretical physics which Hilbert (with his school) had fostered since the beginning of the twentieth century. In Hilbert's wake contributions were given to the resolution of Problem 6 in thermodynamics by Max Born and Constantin Carathéodory, in quantum mechanics by Hermann Weyl and John von Neumann, in electrodynamics by Hermann Minkowski, in probability theory by Richard von Mises and by Sergei Natanovich Bernstein and Andrey Nikolaevich Kolmogorov in the context of the modern measurement theory.

In Hilbert's conception, all sciences, starting from mechanics, should be treated according to the model set forth in geometry. The lecture course held by Hilbert at the University of Göttingen in 1905 on the axiomatization of physical theories (*Logische Principien des mathematischen Denkens*, 1905; see Corry, 2004) showed how this should be done in practice. A section was dedicated to each of the following disciplines: mechanics, thermodynamics, probability calculus, kinetic theory of gases, insurance mathematics, electrodynamics, and psychophysics. (Other sections, on radiation theory and on relativity, were added in the lecture course held on the same subject in 1913.) Hilbert reviewed the various theories of physics and the different disciplines, with no specific references to the historical background or sources, in order to give a unified view of them to the students. In this perspective, it was assumed that every theory was governed by specific axioms which expressed the mathematical properties establishing relationships among the basic magnitudes pertaining to it. In parallel, there were some general mathematical and variational principles assumed to be valid for all physical theories: Great importance was given to the continuity axiom, for which a general formulation was proposed along with other ones more specific to each theory. The aim was to show how, through these (specific and general) principles, one could derive the basic equations of each theory. The derivation of the equations from the axioms was indicated in a schematic, very rapid way; Hilbert often confined himself to affirming the possibility of such a derivation, considered as plausible and feasible according to the pattern shown for geometry in *Grundlagen der Geometrie*. There was furthermore no general demonstration of the independence, consistency or completeness of the axiomatic systems proposed.

According to this scheme, as mentioned earlier, Hilbert treated psychophysics as the last item after the other disciplines listed. With reference to the work of the Austrian astronomer Egon Ritter von Oppolzer (1869-1907) on the psychological theory of color perception (Oppolzer, 1902-1903), which was based on Fechner's law, the idea was to express the magnitude of brightness (a purely psychological parameter) as a function of the intensity and wavelength (physical parameters).

The axioms assumed to be defined for a collection of "brightnesses" x_1, x_2, …, x_n, as confirmed, according to Hilbert, by experience, were the following (Corry, 2004):

1. *To every pair of brightnesses x_1, x_2, a third one $[x_1,x_2]$ can be associated, called "the brightness of the mixed light of x_1, x_2." Given a second pair of brightnesses x_3, x_4, such that $x_1 = x_3$ and $x_2 = x_4$, then $[x_1,x_2] = [x_3,x_4]$.*

2. *The "mixing" of various brightnesses is associative and commutative.*

3. *By mixing various homogeneous lights of equal wavelengths, the brightness of the mixed light has the same wavelength, while the intensity of the mixed light is the sum of the intensities.*

Let us indicate $[x_1,x_2]$ with $x_{12:}$ We can write x_{12} as a function of the two parameters x_1, x_2, so that

$$x_{12} = f(x_1,x_2).$$

We can then introduce a function F such that

$$F(x_{12}) = F(f(x_1,x_2)) = F(x_1) + F(x_2).$$

From Axiom 3 and from the general postulate of continuity, F, for homogeneous light, is proportional to the intensity. According to Hilbert, the knowledge of this function, which was called "stimulus value", allowed one to establish the whole theory.

Hilbert's work, which is very short, schematic, without reference to the historical background of psychophysics, remains a puzzle from many points of view, particularly concerning Hilbert's relationship with the development of the discipline itself in Germany, and especially in Göttingen, an important driving force of the psychology of the time (see also Corry, 2004).

Acknowledgements

This research is co-financed by FSE, Regione Lombardia.

References

Algom, D. (2011). Preface. In D. Algom, D. Zakay, E. Chajut, S. Shaki, Y. Mama, & V. Shakuf (Eds.), *Fechner Day 2011. Proceedings of the 27th Annual Meeting of the International Society for Psychophysics* (pp. vii-viii). Raanana, Israel: The International Society for Psychophysics.

Brožek, J., & Gundlach, H. (Eds.) (1988). *G. T. Fechner and psychology*. Passauer Schriften zur Psychologiegeschichte, Nr. 6. Passau: Passavia Universitätsverlag.

Corry, L. (1997). David Hilbert and the axiomatization of physics (1894-1905). *Archive for History of Exact Sciences*, *51*, 83-198.

Corry, L. (2004). *David Hilbert and the axiomatization of physics (1898-1918). From Grundlagen der Geometrie to Grundlagen der Physik*. Dordrecht et al.: Kluwer.

Fechner, G. T. (1860). *Elemente der Psychophysik*. Leipzig: Breitkopf & Härtel [Reprint Amsterdam: Bonset, 1964].

Fechner, G. T. (1877). *In Sachen der Psychophysik*. Leipzig: Breitkopf & Härtel [Reprint Amsterdam: Bonset, 1968].

Fechner, G. T. (1882). *Revision der Hauptpuncte der Psychophysik*. Leipzig: Breitkopf & Härtel [Reprint Amsterdam: Bonset, 1965].

Heidelberger, M. (1993). *Die innere Seite der Natur: Gustav Theodor Fechners wissenschaftlich-philosophische Weltauffassung*. Frankfurt am Main: Klostermann [English edition *Nature from within: Gustav Theodor Fechner and his psychophysical worldview* by C. Klohr. Pittsburgh: University of Pittsburgh Press, 2004].

Hilbert, D. (1899). *Grundlagen der Geometrie*. Leipzig: Teubner [10. edition, 1968].

Hilbert, D. (1905). *Logische Principien des mathematischen Denkens* (HLN SS 1905, annotated by E. Hellinger). Quoted in L. Corry (2004).

Hilbert, D. (1918). Axiomatisches Denken. *Mathematische Annalen*, *78*, 405-415.

Klev, A. (2011). Dedekind and Hilbert on the foundations of the deductive sciences. *The Review of Symbolic Logic*, *4*(4), 645-681.

Kline, M. (1972). *Mathematical thought from ancient to modern times*. Oxford: Oxford University Press.

Leonard, R. (2010). *Von Neumann, Morgenstern, and the creation of game theory*. Cambridge et al.: Cambridge University Press.

Murray, D. J. (1993). A perspective for viewing the history of psychophysics. *Behavioral and Brain Sciences*, *16*, 115-186.

Oppolzer, E. Ritter von (1902-1903). Grundzüge einer Farbentheorie. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, *29*, 183-203; *33*, 321-354.

Pyenson, L. (1979). Mathematics, education, and the Göttingen approach to physical reality, 1890-1914. *Europa, A Journal of Interdisciplinary Studies*, *2*(2), 91-127.

Rowe, D. E. (1989). Klein, Hilbert, and the Göttingen mathematical tradition. In K. M. Olesko (Ed.), *Science in Germany. The intersection of institutional and intellectual issues* (pp. 186-213). *Osiris*, 2. Series, Volume 5.

Zudini, V. (2009). *I numeri della mente. Sulla storia della misura in psicologia*. Trieste: EUT.

Zudini, V. (2011). The Euclidean model of measurement in Fechner's psychophysics. *Journal of the History of the Behavioral Sciences*, *47*(1), 70-87.

THE “LAUNCHING EFFECT” DEPENDS ON SIZE OF COLLIDING OBJECTS

Michele Vicovaro, Luigi Burigana
Department of General Psychology, University of Padua, Italy
michele.vicovaro@studenti.unipd.it luigi.burigana@unipd.it

Abstract

Suppose that two squares are aligned horizontally on a computer screen. At a point in time one square (A) starts moving towards the other (B), which remains stationary. When A and B come into contact, the latter starts moving with the same velocity as A, while A comes to a stop. Albert Michotte (1946) demonstrated that, under these stimulus conditions, observers usually have the impression that A “launches” or “pushes” B, a perceptual effect he called the “Launching Effect”. Michotte claimed that features of objects A and B (e.g., their size) exert only a slight influence on the phenomenon, however to our knowledge no research has systematically tested this claim. In our experiment we manipulated the size of two simulated spheres, and found that the velocity ratio most favouring the perception of the Launching Effect actually depends on the size of both spheres.

Many philosophers and psychologists agree that the concept of causality is the “cement” of the universe. As soon as one grasps the concept of causality, events in the world start to be conceived as relations between causes and effects. This makes an evolutionary advantage: cause-effect connections may be generalized in order to predict future events. The Belgian researcher Albert Michotte (1946) was a pioneer in demonstrating that causality can be directly perceived. He performed novel experiments showing how the perception of causality directly depends on stimulus conditions. In one of his experiments, he presented observers with two small squares aligned horizontally. At a moment one square (*A*) started moving towards the other (*B*), which remained stationary. When *A* and *B* came into contact, the latter started moving with the same velocity as *A*, while *A* came to a stop (see Figure 1 for a 3D version of Michotte’s stimuli). Observers reported the impression that *A* “launched” or “pushed” *B*, a perceptual effect called the “Launching Effect”. Michotte maintained that the Launching Effect is a purely perceptual phenomenon, that it may be explained in terms of Gestalt-theoretic principles, and is not influenced by observers’ experience with real mechanical collisions. Scholl and Tremoulet (2000) reformulated Michotte’s theorization in terms of perceptual module, thus emphasizing the “cognitive impenetrability” of the phenomenon. A compelling argument for this thesis is that the Launching Effect occurs even when *A* and *B* are fuzzy coloured shadows (Michotte’s Experiment 27), or when object *A* is a real wooden sphere, and object *B* is just a shadow (Michotte’s Experiment 28). This suggests that physical plausibility of collisions is not a necessary condition for the perception of the phenomenon, thus discrediting the possible role of past experience with real collisions. Recent experimental findings have further supported this claim: given appropriate contextual stimuli, the Launching Effect may occur even when spatiotemporal relations between *A* and *B* should exclude real collisions (Choi & Scholl, 2004; Bae & Flombaum, 2010).

In addition, Michotte (1946, p. 78) reported the qualitative observation that shape and dimension of *A* and *B* may influence the “vividness” of the Launching Effect, but classified this influence as slight and marginal. He did not elaborate this qualitative observation in detail. In partial contrast with Michotte’s view of “purely perceptual” character of the effect we are discussing, the possible influence of objects’ shape and dimension on its perception might suggest that observers' experience with real collisions can have a role in the process. Despite its importance, this topic has not received much attention in the literature on the perception of causality. In an early study, Natsoulas (1961) presented observers with classic Michottean stimuli, and varied both velocity and size of the stimuli. His results confirmed that size had a relatively small influence on the Launching Effect. However, one possible reason for this finding could be that Natsoulas used large steps of variation of the velocity ratio between *A* and *B*, and this may have overshadowed the effect of size.

People without formal instruction in Physics correctly predict that postcollision velocity of object *B* is decreasing with the size (implied mass) of *B* and decreasing with the size of *A* (De Sá Teixeira, De Oliveira, & Viegas, 2008; Vicovaro, in press). Even 5.5-6.5 months old infants expect the outcome of a collision to be dependent on size of colliding objects (Kotovski & Ballairgeon, 1998). The aim of our experiment is to verify whether the size of colliding objects also influences visual perception of the Launching Effect. Michotte's claim that perception of the Launching Effect is not influenced by past experience with mechanical collisions would be supported if we find that size of the stimuli does not influence perception of the Launching Effect. Otherwise, we should conclude that observers' experience with mechanical events may exert at least a partial influence on the phenomenon.

Experiment

One critical variable for the perception of the Launching Effect is the ratio between the pre-collision velocity of object *A* ($v(A)$) and the post-collision velocity of object *B* ($v(B)$). In particular, Michotte (1946) reported that the Launching Effect is replaced by a "Triggering Effect" when $v(B)$ is twice $v(A)$. *Triggering Effect* means that the post-collision motion of *B* appears self-generated, rather than generated by the collision with *A*. Natsoulas (1961) also found that when $v(A)$ is three times $v(B)$, observers have the impression of "braked launch", i.e., the impression that the post-collision motion of *B* is braked by some force, rather than exclusively generated by the collision with *A*. Even though the *Braking Effect* is not reported in most studies on perception of causality, its existence was proved by Minguzzi (1968) in an extensive series of experiments. In our experiment we presented observers with a three-dimensional version of Michotte's stimuli (see Figure 1), and determined a "braking threshold" and a "triggering threshold". The *braking threshold* is the minimum value of ratio $v(A)/v(B)$ above which observers will perceive the Braking Effect. The *triggering threshold* is the minimum value of ratio $v(B)/v(A)$ above which observers will perceive the Triggering Effect. When ratios $v(A)/v(B)$ and $v(B)/v(A)$ are both below their respective braking and triggering thresholds, then observers will perceive the Launching Effect.

In our experiment, we wanted to determine whether the braking and triggering thresholds depend on size of colliding objects. If it is true that experience with physical events of collisions influences the Launching Effect, then when *A* is larger than *B* (and heavier, for objects of equal material) observers should perceive the Launching Effect with relatively high $v(B)/v(A)$ ratios, because this is what it would happen in real physical collisions. For the very same reason, observers should not perceive the Launching Effect with high $v(A)/v(B)$ ratios. In terms of braking and triggering thresholds, we should find a high triggering threshold but a low braking threshold. The opposite should be true when the size of *A* is smaller than the size of *B*: in this case we predict a high braking threshold and a low triggering threshold.

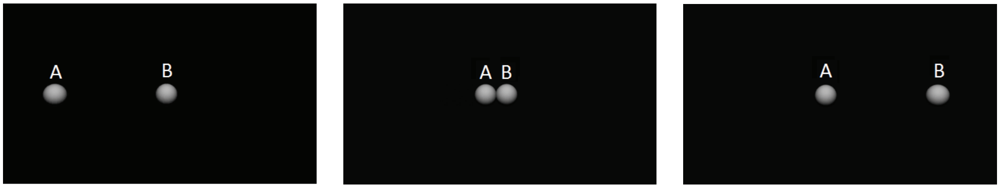


Figure 1: Three frames of an animation sequence used in our experiment (a 3-D version of Michotte's stimuli). Labels "*A*" and "*B*" are added for reference in our discussion.

Participants. Fifteen students of Psychology (aged from 20 to 29, 4 males) participated in the experiment. They all had normal or corrected-to-normal visual abilities, and were paid for the participation.

Stimuli and apparatus. The stimuli were presented on a personal computer equipped with a 37.5 cm × 30 cm screen and a keyboard. Participants sat at a distance of about 50 cm from the screen, the background of which was black. Two smooth, greenish 3-D spheres (created by 3D Studio Max) were presented at middle height of the screen. At the beginning of each animation, one sphere (*A*) appeared close to the left edge of the screen and the other sphere (*B*) in the centre. Then, 170 milliseconds after the appearance of the spheres, *A* began to move horizontally from left to right towards *B*, until making contact with it. At this point, *A* came to a stop, and *B* started moving in the same direction as *A*, until stopping close to the right edge of the screen (see Figure 1). We manipulated the apparent size of *A* and *B*, according to a 3 size *A* (4.2, 8.4, 16.8 cm^3) × 3 size *B* (4.2, 8.4, 16.8 cm^3) factorial design. These sizes (volumes) of the spheres are computed on the diameters of the corresponding images on the screen. The velocity of *A* was kept the same (15.5 cm/s) across the experiment. In each of the nine experimental conditions we manipulated the velocity of *B* for determining the *braking* and *triggering* thresholds (see next paragraph).

Procedure and experimental design. Instructions readable on the screen informed the participants that they would be presented with two colliding spheres, which could represent a collision, e.g., between billiard balls. Participants were asked to pay attention to the post-collision velocity of the initially stationary sphere (*B*), and judge whether its motion was "natural" or "unnatural" compared with the force exerted by the initially moving sphere (*A*). The instructions specified that "unnatural" could have two alternative meanings: first, that the motion of *B* was too slow compared with the force exerted by *A*, as if the motion of *B* was braked by an invisible force; second, that the motion of *B* was too fast compared with the force exerted by *A*, as if the motion of *B* was accelerated by an invisible force. In each trial participants were allowed to view the stimulus as many times as they wanted by pressing "SPACE" on the keyboard and then, when they felt ready to respond, they had to press "N" for the "natural" response, and "Z" for the "unnatural" response. After the instructions, participants were presented with five randomly chosen stimuli to familiarize with the task. They were recommended to rely on their visual impression, not on what they knew from experience or from learning of Physics.

In order to estimate the 50% braking and triggering thresholds we used the standard psychophysical method of "randomly interleaved staircases" with fixed step size (Levitt, 1971). For the estimation of the braking threshold we manipulated the velocity of *B* (the velocity of *A* was fixed at 15.5 cm/s) such that the $v(A)/v(B)$ ratio could take on 11 values from 1 to 3 in steps of 0.2. The threshold was estimated by generating two staircases, one ascending and the other descending. The *ascending staircase* started from the lowest value, corresponding to a velocity ratio of 1, which gave rise to a clear launching impression. Every time the participant responded “natural”, the velocity ratio was increased by one step (for instance, from 1 to 1.2, then to 1.4, etc.), until the participant responded “unnatural” (she perceived a braked launch). At that point, the staircase changed its direction, and the velocity ratio was decreased by one step every time the participant responded “unnatural” (for instance, from 2 to 1.8, then to 1.6, etc.). The staircase changed its direction whenever the participant changed her answer, and continued in that direction until the participant changed her answer again. The series of stimuli between two changes of direction is called a “run”. Symmetrically, the *descending staircase* started from the highest value, corresponding to a velocity ratio of 3. The velocity ratio was decreased as long as the participant responded “unnatural” (she perceived a braked launch), and the staircase changed its direction when the participant changed her response. Both staircases were terminated after eight runs and the 50% braking threshold was estimated by averaging the midpoints of the last four runs of both staircases (Levitt, 1971, p. 470). For the estimation of the triggering threshold we applied the same procedure, only referring to ratio of velocities $v(B)/v(A)$ (rather than $v(A)/v(B)$).

Both braking and triggering thresholds were estimated in each of the 9 (3 Size A × 3 Size B) experimental conditions. In order to avoid anticipatory effects, the 36 staircases (9 experimental conditions × 2 thresholds × 2 staircases) were randomly interleaved. Participants were allowed to rest as much as they wanted after every 200 trials. The experimental session could last from 35 to 45 minutes.

Results

Parts (i) and (ii) of Figure 2 show the means (over participants) of the braking and triggering individual thresholds, for each condition of our experimental design (i.e., each combination of size of spheres A and B). For graphical illustration and statistical analysis we find it convenient to represent the nine conditions as the levels of one factor (combining variables “size A” and “size B”), which we call “size of the stimuli”. The labels of the nine levels of this combined factor are shown on the horizontal axis of parts (i) and (ii) of the figure (see also caption to the figure). Labels “S”, “M”, and “L” on the abscissa stand for small, medium, and large size of each sphere (4.2, 8.4, 16.8 cm^3) .

Braking threshold. A one-way within subjects ANOVA showed that factor “size of the stimuli” has a significant effect on the braking threshold ($MS_e = 0.852$, $F(8,112) = 6.073$, $p = 1.76 \times 10^{-6}$). As shown in Figure 2.i, the braking threshold tends to be higher when A is smaller than B, and lower when A is larger than B. In other words, with large $v(A)/v(B)$ ratios, observers tend to see more easily a “braked launch” when A is larger than B, whereas they tend to see more easily a “natural launch” when A is smaller than B. This result is in agreement with physical principles, and supports the hypothesis that size of colliding objects does influence perception of the Launching Effect. However, caution is required in interpreting these data. A Tukey’s post-hoc test showed that there are only three significant differences between the nine experimental conditions: in condition (L,S) the braking threshold is significantly lower than in conditions (S,L) ($p < 0.05$), (M,L) ($p < 0.05$), and (S,M) ($p < 0.05$). In other words, only extreme differences in size produce significantly different braking thresholds. Note that for the equal-size conditions ((S,S), (M,M), (L,L)), the braking threshold is higher than expected (about 2.5). This means that when A is the same size of B, and $v(A)/v(B)$ ratio is about 2.5, observers still perceive a “natural launch” 50% of the time. Contrary to the results reported by Natsoulas (1961), our results suggest that only very large $v(A)/v(B)$ ratios can produce a vivid Braking Effect. It is thus possible that our manipulations

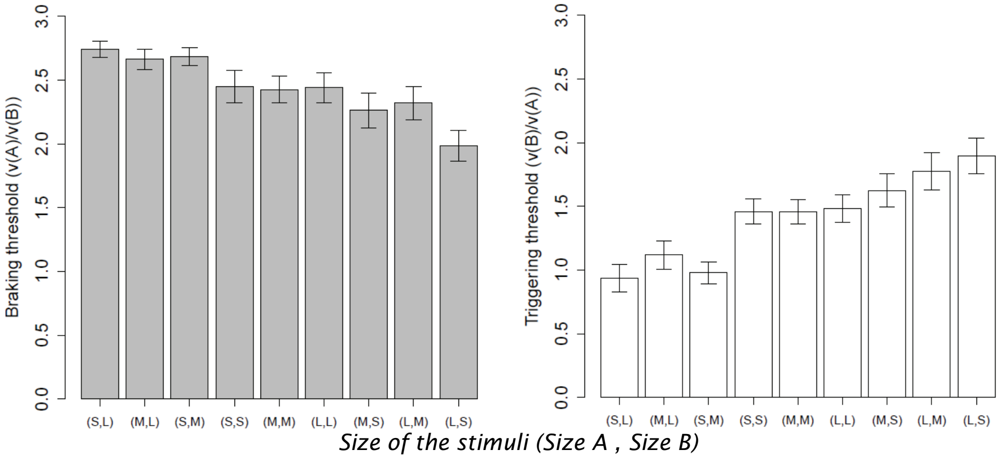


Figure 2. (i) (left): Mean *braking* threshold for each experimental condition. (ii) (right): Mean *triggering* threshold for each experimental condition. Labels "*S*", "*M*", and "*L*" on the abscissa stand for small, medium, and large size of each sphere (4.2, 8.4, 16.8 cm^3). We ordered the nine levels on the abscissa trying to obtain approximately monotone trends.

of velocity ratio were not large enough to produce clear Braking Effects, and that larger manipulations would bring out stronger effects of size on the braking threshold.

Triggering threshold. A one-way within subjects ANOVA showed that factor "size of the stimuli" has a significant effect on the triggering threshold ($MS_e = 1.734$, $F(8,112) = 11.46$, $p = 8.97 \times 10^{-12}$). As shown in Figure 2.ii, the triggering threshold tends to be higher when *A* is larger than *B*, and lower when *A* is smaller than *B*. In other words, with large $v(B)/v(A)$ ratios, observers tend to see more easily a Triggering Effect when *A* is smaller than *B*, whereas they tend to see more easily a "natural launch" when *A* is larger than *B*. Tukey's post-hoc test showed that in all the three conditions where *A* is smaller than *B* ((*S*,*L*), (*M*,*L*), (*S*,*M*)) the triggering threshold is significantly (or marginally) lower compared with the three conditions where *A* is larger than *B* ((*M*,*S*), (*L*,*M*), (*L*,*S*)). These results are in agreement with physical principles, and support the hypothesis that size of the stimuli influence the triggering threshold.

Discussion

The results of our experiment lend support to the hypothesis that size of *A* and *B* influences the perception of the Launching Effect. This evidence was also reported by Michotte (1946) who however classified it as slight and marginal. Our results suggest that this judgment needs revision. Here we remark two interesting results showing that the influence of size on the Launching Effect is not thus negligible: (1) As expected, when the size of *A* is no smaller than the size of *B* and $v(A) = v(B)$, observers report unambiguous Launching Effect (this finding has been replicated several times by different researchers). However, when *A* is small (4.2 cm^3) and *B* is large (16.8 cm^3) the same velocity ratio produces the Triggering Effect about 50% of the time (see Figure 2.ii). (2) As expected, when the size of *A* is no larger than the size of *B* and $v(B) = 2v(A)$, observers report an unambiguous Triggering Effect (this finding too has been replicated several times). However, when *A* is large and *B* is small the same velocity ratio produces the Launching Effect about 50% of the time (see Figure 2.ii).

To sum up, the relations between size (implied mass) and velocity of the colliding objects most favouring visual perception of the Launching Effect appear similar to those governing physical collisions. This lends support to the idea that observers' past experience with collisions may exert some influence on the perception of the launching phenomenon. Such conclusion seems to us to deviate from Michotte's interpretation of the Launching Effect simply in terms of Gestalt principles, and from Scholl and Tremoulet's (2000) conception of the phenomenon as a perceptual module. Our findings are rather in agreement with White's template-matching hypothesis, according to which visual causal impressions result from the matching of stimulus features to features of templates of causal mechanisms (White, 2005, p. 403). Such templates are presumed to be stored in memory, and contain salient features of causal interactions that allow the visual system to recognize causal events. Note that they need not be entirely isomorphic to physical collisions. Actually, in our experiment sphere *A* remained stationary after collision (an extremely unlikely event in real collisions), and nevertheless the Launching Effect was perceived. The results of our experiment suggest that those templates are characterized by a notable common feature: they provide information concerning objects' size.

References

Bae, G. Y., & Flombaum, J. I. (2010). Amodal causal capture in the tunnel effect. *Perception*, *40*, 74-90.

Choi, H., & Scholl, B. J. (2004). Effects of grouping and attention on the perception of causality. *Perception & Psychophysics*, *66*, 926-942.

De Sá Teixeira, N. A., De Oliveira, A. M., & Viegas, R. (2008). Functional approach to the integration of kinematic and dynamic variables in causal perception: Is there a link between phenomenology and behavioral responses? *Japanese Psychological Research*, *50*, 232-241.

Kotovski, L., & Baillargeon, R. (1998). The development of calibration-based reasoning about collision events in young infants. *Cognition*, *67*, 311-351.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, *49*, 467–477.

Michotte, A. (1946). *La Perception de la Causalité*. Louvain: Institut Supérieur de Philosophie.

Minguzzi, G. (1968). Sulla validità della distinzione tra percezione di nessi causali e percezione di dipendenze funzionali [On the soundness of the distinction between perceiving causal connection and perceiving functional dependence]. In G. Kanizsa & G. Vicario (Eds.) *Ricerche sperimentali sulla percezione [Experimental research on perception]* (pp. 161-196). Trieste: Istituto di Psicologia dell'Università di Trieste.

Natsoulas, T. (1961). Principles of momentum and kinetic energy in the perception of causality. *American Journal of Psychology*, *74*, 394-402.

Scholl, B. J., & Tremoulet, P. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*, 299-309.

Vicovaro, M. (in press). Intuitive physics of collision effects on simulated spheres differing in size, velocity, and material. *Psicológica.*

White, P. A. (2005). Visual causal impressions in the perception of several moving objects. *Visual Cognition*, *12*, 395-404.

PERCEPTION IS NOTHING WITHOUT CONTROL (OF VELOCITY)

Bastianelli A., Vicentini M., Scandola M., Fiorini P.
Altair Laboratory, Computer Science Department, University of Verona, Italy
alessia.bastianelli@ metropolis.sci.univr.it

The integration of haptic and visual perception in a virtual environment is a critical issue when there are delays between haptic and visual rendering. We suppose that integrated perception is affected of both the visual haptic discrepancies and the contact velocity of a target with an object. To test this hypothesis, we performed two experiments in which participants had to actively contact a virtual wall by using a haptic device. In the first experiment, the target velocity was held by the user (no-constant velocity task); in the second, the target velocity was kept constant at about 17mm/s (constant velocity task).We use the method of constant stimuli manipulating the delay of the haptic and visual rendering. Participants were asked to judge if they first perceived the visual or the haptic rendering when they contacted the virtual wall of 331N/m or 83N/m stiffness. The results show that performance in the no-constant velocity task is strongly affected by the stiffness, whereas in the constant velocity task, participants seemed to perceive delays correctly.

One of the main advantages of telepresence systems is they enable individuals to work in remote locations. Network technologies offer the manipulation of real or virtual objects using a haptic interface. A suitable manipulation requires accurate sensor information be provided to the operator, preferably through multiple senses (i.e. visual, haptic, and audio). The human perceptual system automatically integrates information available in all sense modalities (Varadharajan, Klatzky, Unger, Swendsen, & Hollis, 2008). However, network communication-induced artifacts occur in long-distance teleoperation, such as time delays and packet loss (Hirche& Buss, 2007).

This communication latency can sensibly affect the subjective experience of “presence” and task performance (Kaber, Riley, Zhou, & Draper, 2000; MacKenzie& Ware, 1993). This issue has received significant research interests in many fields, including human computer interaction and virtual reality, and results are not always unambiguous.

Several studies on the influence of delayed feedback have examined haptic and visual perception delay (Alhalabi, Horiguchi, &Kunifuji, 2003; MacKenzie& Ware, 1993; Lane et al., 2002; Kaber et al., 2000). Delays can modify a user’s perception of virtual objects (i.e. perception of stiffness). In particular, Wu, Basdogan, and Srinivasan (1999) showed that stiffness perception is influenced by visual information. A recent study by Jay, Glencross and Hubbold (2007) examined the impact of delayed haptic and visual feedback. They found both visual and haptic delay retarded task performance in terms of difficulty of contact with the target. In contrast to the findings of Wu et al. (1999), they asserted that haptic delay had a greater impact on performance than visual latency.

Recently, Gleeson and Provancher (2011) found that the contact velocity of a target with an object did not have a significant influence on the perception of stiffness. Nevertheless, according to Wu, Abbott, and Okamura (2005) the velocity of the target played an important role: they found a better precision with low-level velocity. Similarly, Vicentini and Botturi (2008) underlined the relationship between contact velocity and stiffness on threshold values. Thus, in general, detection thresholds for visual and haptic delays may vary substantially (Adelstein et al., 2003; Allison, Harris, Jenkin, Jasiobedzka, &Zacher, 2001; Jay et al., 2007), probably depending on both the visual haptic discrepancies and the contact velocity of a target with an object.

Within this framework, we designed two experiments in which we proposed for the user two virtual surfaces. The first surface was rendered through the visual channel. The second one was experienced by the user through a force feedback, simulating a pliable surface. This was spatially separated by the first surface by a variable temporal delay.

Methods

We tested participants' perception in two multimodal experiments with temporal discrepancies between the visual and the haptic virtual walls (the visual wall was not haptically rendered, and the haptic wall was not visually rendered). The movement of the tool was connected in the visual scene to a small red sphere which acted as a proxy for the position of the tool tip in the virtual world. The task was to contact a virtual wall and immediately retract. To ease the haptic perception and to make the participants' movements easily controllable across the whole trial, we opted for a displacement along the z-axis (near-far horizontally). Thus, each subject was instructed to keep the elbow on a horizontal surface. Movements along the other directions were neglected. Furthermore, to ease the visual perception we opted for a 2D scenario in which the red sphere moves along the y-axis (bottom-up, vertically). Basically, the sphere was seen "from above" and we map the near-far arm movements into a bottom-up sphere movement. In contrast, a 3D visualization where the sphere movement was coherent with the arm movement (near-far) would add possible depth perception problems or issues with movement along the x-axis (for example, the user could have moved the red sphere laterally). After each trial participants were asked to indicate which wall they first encountered. The response was done via key press and the data were logged.

Participant.A total of seven subjects were examined for Experiment I, and a different sample of seven subjects took part in Experiment II (age ranged from twenty to thirty y.o. all male and well experienced with haptic devices). Allparticipants were recruited within the laboratory staff. They were not informed about the experiment goals and were simply instructed how to attend to the task. All participants had normal or corrected-to-normal vision and had a normal sense of touch. They used their dominant hands to perform the tasks. All of them gave consent for the use of personal data for the purposes of this scientific research.

Apparatus.To simulate realistic force feedback, we used a Freedom 7S force feedback haptic device (MPB Technologies, Montreal, Quebec), which provides a workspace of about 170 W x 220 H x 330 D mm, a position resolution of 2µm and a resolution in force rendering of 40mN, with a maximum update rate above 1kHz. The pen-hold grasping configuration involved the thumb, index, and middle fingers. For the visual rendering, we used a 22-inch widescreen monitor, placed in front of the subject at a distance of about 50cm. The visual scene for our experiment was generated using the OpenGL library and rendered on the monitor. The force feedback returned by the haptic device was generated by a custom C++ program based on the provided Freedom API. The running OS was Windows 7 for the first experiment and Ubuntu 9.04 for the second.

***Experiment 1** – No constant velocity*. This experiment was aimed to detect the perception of temporal discrepancies between a haptic and visual surface in a condition of no constant velocity. The participants' tool velocity was not kept constant and they could use the velocity they preferred. The temporal delays were ±1250, 650, and 250ms. The stiffness values for the haptic wall were 331N/m and 83N/m, respectively the human skin and fat stiffness values (Gerovich, Marayong, & Okamura, 2004). The experimental design was a 2x6 within-subjects design.

***Experiment 2** - Constant velocity.*This experiment was aimed to determine if perception of delays between a haptic and a visual surface is influenced by the use of a constant velocity. For this reason, the subjects were asked to move at the same velocity as a yellow sphere (17 mm/s, optimal velocity to discriminate different stiffness, as shown in Vicentini &Butturi, 2010). In this experimental session, the stiffness factor was equal to the previous one, while temporal delays were ±1470, 882, and 294ms.

Statistical Analysis. Statistical analyses were conducted separately for each subject, each experiment, and for aggregate data using the R framework (R Core Development Team, 2011). Log-linear analysis was used to determine whether there were statistical significant differences due to stiffness, delay, or experimental session. For this purpose, delays were grouped in six categories:

First Haptic (FH) high (-1470 and -1250ms), FH medium (-882 and -650 ms), FH low (-294 and -250ms), First Vision (FV) low (250 and 294ms), FV medium (650 and 892ms), FV high (1250 and 1470ms). In log-linear analysis we considered the subject factor such as a random factor.

In addition, we fitted to subjective responses psychometric logistic functions using theNonlinear Least Squares algorithm *nlm* and computed PSEs and DLs.

Results

Psychophysical curves, PSEs, and DLs values for Experiment 1 and 2 are reported respectively in Figure 1 and Table 1.For experiment 1 the average contact velocity was 29.79 mm/s. Log-linear analysis was computed on the best fitting model, selected considering the lowest Bayesian information criteria (BIC) on the seven models reported below.

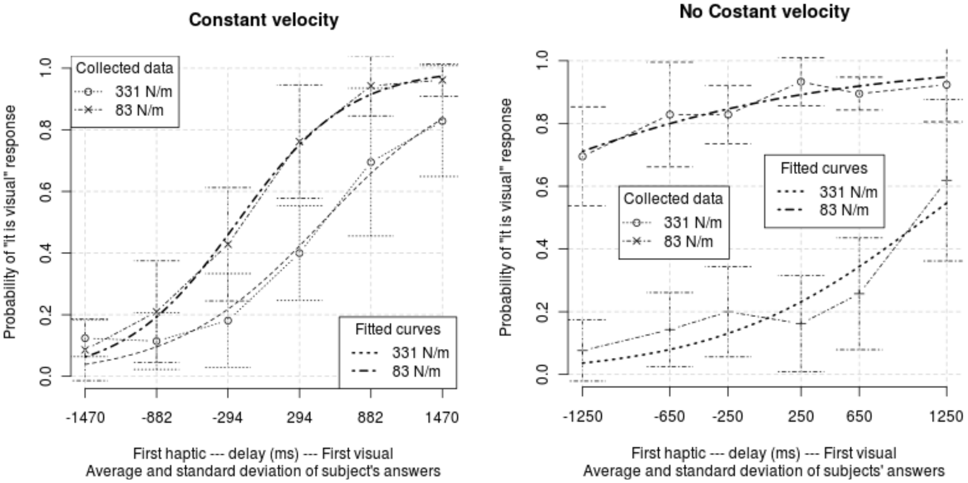


Figure 1: Psychophysical curves for constant and no constant velocity experiments

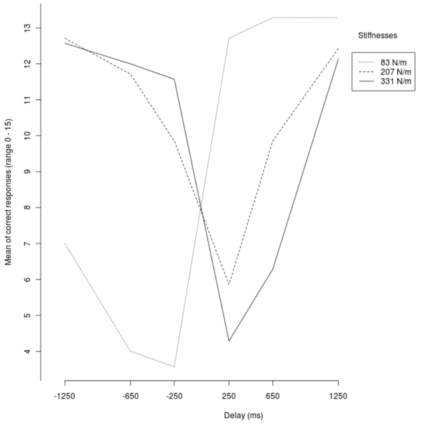


Figure 2: Interaction plot *stiffness:experiment*s. In the y-axis the mean of subjective correct responses, in the x-axis the stiffness

In the saturated model (1) we used as fixed factors *delay*, *stiffness*, *experiment* and their 2-way and 3-way interactions, and as random factor *(1|subject)*. BIC criteria for each model are reported in Table 2. The resulting best fitting model is (5) and a graphical representation for *stiffness:experiments*interaction isreported in Figure 2.Tukey group-to-group comparisons for stiffness and experiment factors showed statistical significant differences (z = -4.401, p<.0001; z = 8.871, p<.0001 respectively). Tukey comparisons for delay factor are reported in Table 3.

$$y = \mathit{delay} + \mathit{stiffness} + \mathit{experiment} + \mathit{delay}{:}\mathit{stiffness} + \mathit{stiffness}{:}\mathit{experiment} + \mathit{delay}{:}\mathit{experiment} + \mathit{delay}{:}\mathit{stiffness}{:}\mathit{experiment} + (1|\mathit{subject}) \tag{1}$$

$$y = \mathit{delay} + \mathit{stiffness} + \mathit{experiment} + \mathit{delay}{:}\mathit{stiffness} + \mathit{stiffness}{:}\mathit{experiment} + \mathit{delay}{:}\mathit{experiment} + (1|\mathit{subject}) \tag{2}$$

$$y = \mathit{delay} + \mathit{stiffness} + \mathit{experiment} + \mathit{delay}{:}\mathit{stiffness} + \mathit{delay}{:}\mathit{experiment} + (1|\mathit{subject}) \tag{3}$$

$$y = \mathit{delay} + \mathit{stiffness} + \mathit{experiment} + \mathit{stiffness}{:}\mathit{experiment} + \mathit{delay}{:}\mathit{experiment} + (1|\mathit{subject}) \tag{4}$$

$$y = \mathit{delay} + \mathit{stiffness} + \mathit{experiment} + \mathit{delay}{:}\mathit{stiffness} + \mathit{stiffness}{:}\mathit{experiment} + (1|\mathit{subject}) \tag{5}$$

$$y = \mathit{delay} + \mathit{stiffness} + \mathit{experiment} + (1|\mathit{subject}) \tag{6}$$

$$y = 1 + (1|\mathit{subject}) \tag{7}$$

	Constant velocity		No constantvelocity	
	331N/m	83 N/m	331 N/m	83 N/m
PSE	-222.73	490.00	1123.74	-1250.00
DL	504.85	653.33	126.26	252.52

Table 1: PSEs and DLs for constant and no constantvelocity experiments

model	df	BIC
Saturated model (1)	25	1348.551
(2)	20	1315.637
(3)	19	1386.926
(4)	15	1518.566
(5)	15	1295.889
(6)	9	1587.594
Null (7)	2	1807.304

Table2: Bayesian Information Criteria

	Estimate	Std. Error	z value	Pr(>|z|)
{-294 -250} = {-1470 -1250}	1.07292	.24732	4.338	< .001
{294 250} = {-1470 -1250}	1.71444	.23179	7.396	< .001
{882 650} = {-1470 -1250}	2.05791	.22659	9.082	< .001
{1470 1250} = {-1470 - 1250}	2.14143	.22549	9.497	< .001
{294 250} = {-882 -650}	1.27766	.19406	6.584	< .001
{882 650} = {-882 -650}	1.62113	.18782	8.631	< .001
{1470 1250} = {-882 -650}	1.70466	.18649	9.141	< .001
{294 250} = {-294 -250}	.64151	.15445	4.154	< .001
{882 650} = {-294 -250}	.98499	.14652	6.722	< .001
{1470 1250} = {-294 -250}	1.06851	.14482	7.378	< .001

Table 3 Tukey group-to-group comparisons for *delay* factor. Statistically significant results p > .001 omitted

Table 4 shows comparisons for the interaction stiffness: experiment. Finally, in Figure 3, the significance Tukey comparisons for delay: stiffness are graphically presented.

	Estimate	Std. Error	z value	Pr(>|z|)
83.Constant_velocity= 331.Constant_velocity	-0.36357	0.08297	-4.382	< 0.001
331. No Constant_velocity = 331.Constant_velocity	0.29611	0.08454	3.502	0.00243
83. No Constant_velocity = 331.Constant_velocity	-1.32832	0.12531	-10.600	< 0.001
331. No Constant_velocity = 83.Constant_velocity	0.65968	0.09168	7.195	< 0.001
83. No Constant_velocity = 83.Constant_velocity	-0.96475	0.13023	-7.408	< 0.001
83. No Constant_velocity = 331. No Constant_velocity	-1.62443	0.11293	-14.384	< 0.001

Table 4: Results for Tukey group-to-group comparisons for the *stiffness:experiment* factor

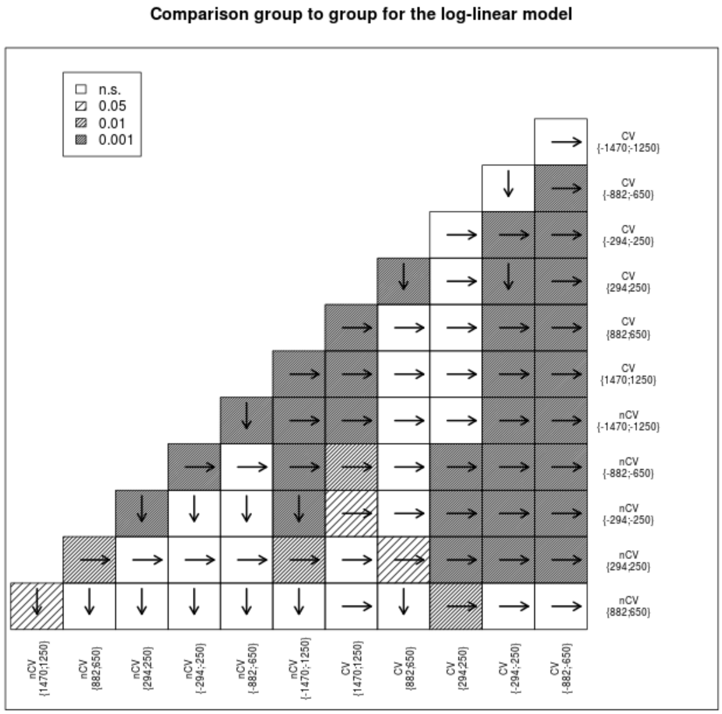


Figure 3: MISCA significance graphical representation for *delay:stiffness*(CS: Constant velocity, nCV: No Constant Velocity)

Conclusions

Based on inspection and statistical analysis, we found that with a constant velocity of 17 mm/sfor both 83N/m and 331N/m stiffness values, participants seemed to perceive delays correctly, although with more accuracy in the stiffest condition. When participants were able to use the velocity they preferred, the results showed that for the low stiffness value, the "it is visual" response is preferred for all delays used; whereas, for the high stiffness value, the trend is inverted, participants seemed to better perceive haptic rendering, except for the extreme visual delay value. In general constant velocity seemed to provide a better performance in perceiving delays; for the no-constant-velocity, which was around 29 mm/s, the responses seemed to be almost completely predicted by the stiffness. This result can be also seen by interaction plots (Fig. 2) which showed that the mean of correct responses in the experiment with constant velocity was high: independent of the stiffness used, the value was always around 0.8. When participants could choose the velocity to move the target only with a high stiffness value they answered correctly. With a low stiffness value such as high stiffness values, their judgments seemed to be very poor.

Globally ouroutcomes do not support Gleeson and Provancher's (2011) results. We found that the velocity of a target seemed to have a significant influence on the perception of stiffness. These data seem to be in agreement with Vicentini and Botturi's results, which indicated a relationship between velocity and stiffness perception.

Further efforts are needed to better understand whether constant velocity per se determines a better detection or whether these results are due to the fact that the velocity of 17mm/s eases the detection, given that it was always lower than the mean velocity used in the experiment with no constant velocity. Additional studies will be carried out in the future to give further insight into the role of contact velocity of a target on haptic and visual perception delay.

ACKNOWLEDGMENTS

This research was supported by SAFROS project (Patient safety in robotic surgery; http://www.safros.eu) which has received funding from Seventh Framework Programme (FP7/2007-2013) under grant agreement n 248960.

References

Adelstein, B. D., Lee, T. G., & Ellis, S. R. (2003). Head tracking latency in virtual environments: psychophysics and a model. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2083–2087.

Alhalabi, M. O., Horiguchi, S., & Kunifuji, S. (2003). An experimental study on the effects of network delay in cooperative shared haptic virtual environment. Computers & Graphics, 27 (2), 205 – 213.

Gerovich, O. Marayong, P., &Okamura, A. M. (2004). "The Effect of Visual and Haptic Feedback on Computer-Assisted Needle Insertion ," Computer-Aided Surgery, 9 (6) 243-249, 2004.

Hirche, S. & Buss, M. (2007). Advances in Telerobotics, chapter 12, Human Perceived Transparency with Time Delay, Springer ,191–209.

Jay, C., Glencross, M., &Hubbold, R. (2007). Modeling the effects of delayed haptic and visual feedback in a collaborative virtual environment. ACM Transactions on Computer-Human Interaction, 14 (2), 8/1–31.

Kaber, D. B., Riley, J., Zhou, R., & Draper, J. V. (2000). Effects of visual interface design, control interface type, and control latency on performance, telepresence, and workload in a teleoperationtask. In 44thAnnual Meeting of the Human Factors and Ergonomics Society, 503–506.

Lane, C. J., Carignan, C. R., Sullivan, B. R., Akin, D. L., Hunt, T., & Cohen, R. (2002). Effects of time delay on telerobotic control of neutral buoyancy vehicles. Proceedings of IEEE International Conference on Robotics and Automation, 3, 2874–2879.

MacKenzie, S. I. & Ware, C. (1993). Lag as a determinant of human performance in interactive systems. In CHI '93: Proceedings of the INTERACT '93 conference on Human factors in computing systems, New York, NY, USA.488–493.

R Development Core Team. (2011). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria.

Varadharajan, V., Klatzky, R., Unger, B., Swendsen, R., & Hollis, R. (2008). Haptic Rendering and Psychophysical Evaluation of a Virtual Three-Dimensional Helical Spring, 57-64.

Vicentini, M. &Botturi, D. (2010). Human Factors in Haptic Perception of Pliable Surfaces, Presence: Teleoperators and Virtual Environments, 18, 6, 478–494.

Wu, W., Basdogan, C., &Srinivasan, M.A. (1999). "The Effect of Perspective on Visual-Haptic Perception of Object Size and Compliance In Virtual Environments", the Winter Annual Meeting of ASME'99, DSC, 67, 19-26.

THE ROLE OF VISUAL HAPTIC DELAY IN TELE-OPERATION PROTOCOLS

Scandola M., Bastianelli A., Vicentini M., and Fiorini P.
Altair Laboratory, Computer Science Department, University of Verona, Verona, Italy
michele.scandola@univr.it

Abstract

In tele-operation protocols developed for robot assisted surgery one of main challenges is the introduction of force-feedback, adding the possibility of manipulating delays between haptic and visual signals. In these conditions it is relevant to understand how the human performance is influenced by visual- haptic delays. In the present research, we study the role of temporal visual-haptic delays between a haptic and a visual rendering using a haptic device in a virtual environment. In a full factorial randomized design participants were asked to judge if they first perceived the haptic or the visual rendering. The delay levels (± 1250, 650, $250ms$) *and the stiffness values of the pliable haptic wall* (83, 207, $331N/m$) *have been randomly presented. Log-linear analysis shows that performance is strongly affected by delay (higher is the delay better is the judgment). Moreover, an interaction between delay and stiffness have been observed. Results will discuss in light of the weighted summation model.*

Teleoperation system permits the individual to control in remote environments (Rosen & Hannaford, 2006). Audio, vision and haptic interaction is needed to enable the human operator to immerse into remote environment and it can aid a wide range of application scenarios, such as robotically mediated surgery (i.e. minimally invasive surgery, laparoscopy).

In recent years, many studies have focused on the added value of haptic feedback for task performance (Scandola, Vicentini, Gasperotti, Zerbato, & Fiorini, 2011; Okamura, 2009), i.e. the visual-haptic discrepancy in virtual reality environments (Scandola, Gasperotti, Vicentini, & Fiorini, 2012). In tele-operation protocols developed for robot assisted surgery haptic signals are sent bidirectionally between the master and the slave, and a global control loop is closed over the communication system. However, the transmission resources for communication networks could be limited and important communication constraints are compelled by communication technology and infrastructure in telepresence applications (Akyildiz, Pompili, & Melodia, 2004). High network traffic may also lead to network congestion and hence large transmission time delays (even over $1,000ms$) between haptic and visual signals and/or packet loss. This can lead to instability of the control system or degrade performance of a force-reflecting teleoperator. In addition this could be also a critical issue for patient safety in tele-operated robot-assisted surgery interventions. Therefore, transmission protocols are of high interest for the haptic modality since the loss of information should be perceptually unperceivable.

From this point of view, to deal the deadband-based haptic data reduction and psychophysics seems to be a needed way. It is shown, that the deadband-based data reduction can lead to high reduction rates. Psychophysical studies indicate that the loss of information induced by the algorithm can be considered unperceivable. Teleoperation protocols with time delay and perceptual-coding in time- delayed are considered (Hirche & Buss, 2007; Vittorias, Kammerl, Hirche, & Steinbach, 2009; Vittorias, Rached, & Hirche, 2010).

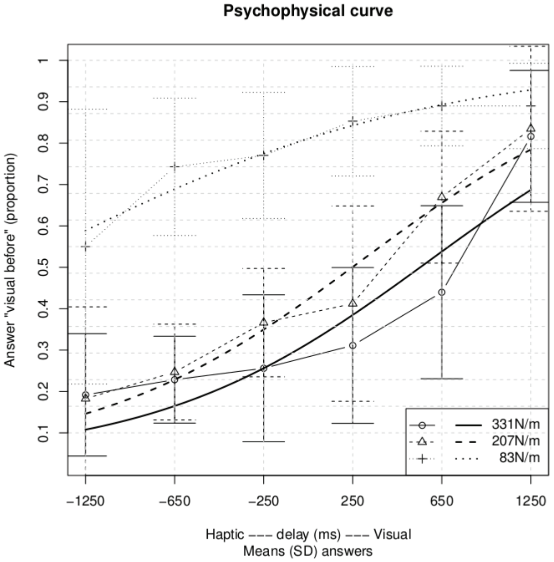


Figure 1: Graphical representation of subjective responses and psychophysical curves fitted on mean values

The weighted summation model (WSM), could be an useful model to predict how users perceive time delays in teleoperation protocols. This model, originally stated by Kuschel, Di Luca, Buss, and Klatzky (2010) but generalizable to different delay conditions (Ley, Haggard, & Yarrow, 2009), postulates that the perceived stimulus is a weighted sum of the effective stimuli as seen in other discrepancy studies (Scandola et al., 2012).

In this work we explore how the human performance is influenced by visual-haptic delays by studying the role of temporal visual-haptic delays between a haptic and a visual rendering. According to WSM, in our experiment Points of Subjective Synchrony (PSS, better described in Section *Statistical Analyses*) have to be influenced by the stiffness of the haptic component, according to the relation that at increasing values of stiffness of the haptic component, correspond a PSS collocation in the continuum of haptic-visual temporal discrepancies, in the direction that starts from the larger haptic delays following a decreasing trend, until the minor visual delays following an increasing trend, and vice versa.

Methods

In order to evaluate the responses to delays between haptic and visual renderings of a stimulus in teleoperation, we designed a Two-Alternative Forced Choice psychophysical experiment based on the method of Constant Stimuli. The experiment was divided in two phases: in the first phase, *”the exploration phase“*, participants had to move along a line in the horizontal plane, away from own body, a haptic tool until they contact a virtual wall (VW). The VW was both haptically and visually rendered, and between the two renderings there was a temporal delay. After each contact with the VW, the participant entered in the second phase, the judgment phase, where s/he had to choose which part of the VW was not delayed, if the haptic or the visual one (that is, they had to answer whether they first perceive the haptic or the visual signal).

We tested six temporal delays (±1250, 650, 250ms) and three haptic stiffness (83, 207, 331N/m) for the haptic rendering. Stiffness levels were selected to simulate pliable objects actually present in the human body (Gerovich, Marayong, & Okamura, 2004): the human fat (83N/m), the human skin (331N/m) and an intermediate stiffness (207N/m).

Participants. Seven subjects took part to the experiment (all right-handed, 2 females, age ranged from 20 to 31 y.o.), with no previous knowledge of the experimental setup. Each experiment lasted about 40 minutes, and no money compensation was planned. All participants had normal or corrected-to-normal vision and without any history of somato-sensory disorders. All participants gave the consent for anonymous use of personal data for the purposes of this scientific research.

Apparatus. Realistic force feedback was rendered using a Freedom 7S force-feedback haptic device (MPB Technologies, Montreal, Quebec). Its workspace can be represented by a parallelepiped about $170mm$ wide, $220mm$ tall and $330mm$ deep. The Freedom is a high performance device, with a position resolution of $2\mu m$, a resolution in force rendering of 40mN and a maximum update rate above 1kHz. The base of this device was positioned so as to be comfortably reached with the subject's dominant hand. The pen-hold grasping configuration involved the thumb, index, and middle fingers. The hand operating the device is not anchored to the desk, hence neither the wrist nor the elbow were provided with a grounded support. For the visual rendering we used a 20-inch wide screen monitor, placed in front of the subject at a distance of about $50cm$.

The implementation of the virtual environment relies on the OpenGL library and on the library provided by the haptic device producer. The VW was graphically rendered in a tridimensional perspective, while the tool tip was connected to a virtual red sphere.

Procedure. We instructed participants saying: "In this virtual environment you have to move the tool close-far along an imaginary line until touching the target VW. The target is graphically and haptically rendered, but between these two components there is a temporal delay. The delayed component randomly varies in each trial. You could enter in contact with just the visual, the haptic component or both components, indifferently. After each trial there will be a black display asking you if you perceived first the haptic or the visual component. Please answer via key press, '1' for the visual component or '2' for the haptic one".

In the *exploration phase* participants had to move a red sphere in direction of the VW, until the contact. The movement of the red sphere was connected with the movement of the haptic tool, and it worked such as a proxy for the position of the tool tip in the virtual world. The VW was rendered in a tridimensional perspective in order not to give cue about the graphical contact point before the contact.

In the *judgment phase* participants had to indicate which was the rendering they firstly encountered in the previous phase. The response was given via key press ('1' for the visual rendering, '2' for the haptic one), and the data were recorded.

Delay and stiffness factors were randomized for each participant, obtaining a 6 × 3 within-subjects design, and every combination was repeated 15 times, for a total of 270 trails. For our analyses, we logged participants' responses.

Statistical Analyses. Statistical Analysis. Statistical analyses were conducted for each subject and for aggregate data using the R framework (R Development Core Team, 2011). Psychometric functions were fitted using the Nonlinear Least Squares algorithm *nlm* over calculated probability of "Visual" response for all stiffness conditions. These functions were defined by the Gauss-Newton logistic function (1):

$$y = \frac{1}{1 + e^{-\beta(x-\psi)}} \tag{1}$$

where the experimental data are y (the subjective response) and x (the delay between

	$331N/m$	$207N/m$	$83N/m$
Mean	762.99	304.83	−1004.69
SD	415.89	520.56	371.34

Table 1: PSS means and SD values

	dof	AIC	BIC
eq. (2)	19	159.1790	213.0684
eq. (3)	9	272.9697	298.4962
eq. (4)	2	295.8183	301.4909

Table 2: AIC and BIC values

visual and haptic stimuli), whereas the parameters to be identified ψ and β are related to the location and the slope of the curve respectively. For each psychometric function the Point of Subjective Synchrony (PSS) was calculated. PSS is the stimulus that elicits 50% of "visual" responses, it is related to the ψ of the psychometric function and it was calculated at proportion 0.5.

Moreover to identify differences in participants' performance we have used the log-linear analysis test, applied to participants' responses, using as fixed factors the delays and the stiffness, and subjects as random factor. We selected the best fitting model among the saturated model, the null model, and the model that considers only the main factors using both Akaiake Information Criterion (AIC) and Bayesian Information Criterion (BIC).

Results and Discussion

Psychophysical curves were fitted to compute PSSs (Table 1). For this analysis we used subjects' responses "it is visual". In Figure 1 are graphically reported the proportions of participants' responses and the fitted psychometric functions on these data.

For the log-linear analysis we selected the best fitting model among the saturated one (2), the null model (4), and a model that considers only the main factors (3).

In Table 2 are reported the AIC and BIC indexes. Following both criteria we selected the best-fitting model, the saturated model (2), to understand which are the factors that could lead to a correct answer.

$$y = \text{delay} + \text{stiffness} + \text{delay} : \text{stiffness} + (1|\text{subject}) \tag{2}$$

$$y = \text{delay} + \text{stiffness} + (1|\text{subject}) \tag{3}$$

$$y = 1 + (1|\text{subject}) \tag{4}$$

where y are the correct subjective responses, *delay* and *stiffness* the main fixed factors, *delay:stiffness* the interaction and $1|subject$ is the random factor.

It is important to highlight that this analysis was done considering the correct responses. It means that for delays where haptic was first, we considered the answers "it is haptic", and where visual was first we considered answers "it is visual", while, previously, to fit psychophysical curves we used answers "it is visual" in all cases.

In Table 4 and 3 results of the significant contrast analysis for the log-linear model Stiffness and Delay main factors have been reported respectively. They show that for the delay factor almost all comparisons reach the statistical significance involve extreme delays. This is probably due to an extreme delayed component which favors the perception of the no-delayed component, while the stiffness factor statistical significant differences are between $83N/m$ and $207N/m$ condition.

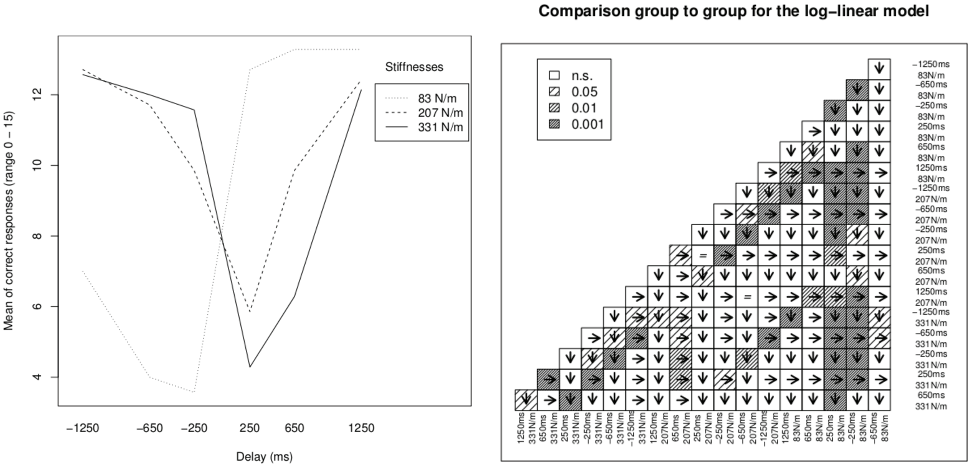
In Figures 2(a) and 2(b) there are two graphical representations for the interaction *delay:stiffness* and statistical significances among cells. Figure 2(a) presents the interaction

	Estimate	SE	z value	p
$-250 = -1250$	$-.337$	.111	-3.048	.0283
$250 = -1250$	$-.418$	.112	-3.748	.00247
$1250 = -650$	.424	.102	4.150	$< .001$
$1250 = -250$	.531	.106	5.017	$< .001$
$1250 = 250$	.613	.107	5.725	$< .001$
$1250 = 650$	.297	.0955	3.108	.0231

Table 3: Results for Tukey group-to-group comparisons for the *delay* factor (no statistical significant results omited)

	Estimate	SE	z value	p
$207 = 83$	.233	.0789	2.947	**.00884**
$331 = 83$	.141	.0818	1.719	.198
$331 = 207$	$-.0920$	.0746	-1.233	.433

Table 4: Results for Tukey group-to-group comparisons for the *stiffness* factor (statistical significant results in bold)



(a) Interaction plot

(b) MISCA representation

Figure 2: (a) Interaction plot for *delay:stiffness* (b) MISCA statistical significance graphical representation for *delay:stiffness*

plot, in Figure 2(b) shows the MISCA plot, where cell represent statistical significance and arrows indicate which group has the larger mean.

From Figure 2(a) it is possible to note the constant tendency in the $83N/m$ condition to underestimate the possibility that the visual component could be delayed, while the influence that $207N/m$ and $331N/m$ have on performance is more complex. Generally, when the haptic component is delayed participants had a correct response (in a similar way to the $83N/m$ condition). This observation could lead us to deduce that the performance is influenced only by the stiffness. However, from both Figure 2(a) and 2(b) it is possible to note that a stiffer haptic component facilitates more correct answers also when the haptic component is strongly delayed.

In conclusion, PSSs are distributed according the WSM, and log-linear analysis seems to show that the model is useful to understand the collocation of PSSs, however it does not predict participants' performance in all cases. Globally our data show that stiffer haptic components permit an easier perception not only of visual-haptic discrepancies, but also the direction of these discrepancies. Further efforts are needed to better understand which factors lead the human perception of visual-haptic discrepancies in a virtual environment and the validity of WSM. In our experiment the visual component of VW is kept constant, avoiding preliminary visual cues about where there will be the graphical contact between the tool and the VW. Further studies will consider different graphical characteristics and how these characteristics influence the human perception.

ACKNOWLEDGMENTS

This research was supported by SAFROS project (Patient safety in robotic surgery; http://www.safros.eu) which has received funding from Seventh Framework Programme (FP7/2007-2013) under grant agreement n 248960.

References

Akyildiz, I. F., Pompili, D., & Melodia, T. (2004). Challenges for efficient communication in underwater acoustic sensor networks. *SIGBED Rev*, *1*(2), 3–8.

Gerovich, O., Marayong, P., & Okamura, A. M. (2004). The effect of visual and haptic feedback on computer-assisted needle insertion. *Computer Aided Surgery*, *9*(6), 243–249.

Hirche, S., & Buss, M. (2007). Transparent data reduction in networked telepresence and teleaction systems. part ii: Time-delayed communication. *Presence: Teleoperators & Virtual Environments*, *16*(5), 532—542.

Kuschel, M., Di Luca, M., Buss, M., & Klatzky, R. L. (2010). Combination and Integration in the Perception of Visual-Haptic Compliance Information. *IEEE Transactions on Haptics*, *3*(4), 234–244.

Ley, I., Haggard, P., & Yarrow, K. (2009). Optimal integration of auditory and vibrotactile information for judgments of temporal order. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(4), 1005-1019.

Okamura, A. M. (2009). Haptic feedback in robot-assisted minimally invasive surgery. *Current Opinion in Urology*, *19*(1), 102–107.

R Development Core Team. (2011). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria.

Rosen, J., & Hannaford, B. (2006). Doc at a distance. *Spectrum, IEEE*, *43*(10), 34–39.

Scandola, M., Gasperotti, L., Vicentini, M., & Fiorini, P. (2012). The role of visual-haptic discrepancy in virtual reality environments. In *Proceedings of the haptic symposium 2012* (pp. 289–295).

Scandola, M., Vicentini, M., Gasperotti, L., Zerbato, D., & Fiorini, P. (2011). Force feedback in psychophysics research: even low performance algorithms may lead to realistic perceptual experience. In *Proceedings of the 27th annual meeting of the international society for psychophysics.*

Vittorias, I., Kammerl, J., Hirche, S., & Steinbach, E. (2009). Perceptual coding of haptic data in time-delayed teleoperation. In *Eurohaptics conference, 2009 and symposium on haptic interfaces for virtual enviroment and teleoperator systmes. worldhaptics 2009* (pp. 208–213).

Vittorias, I., Rached, H. B., & Hirche, S. (2010). Haptic data reduction in multi-dof teleoperation systems. In *Proceedings of the international symposium on haptics audio-visual environments and games (HAVE).*

FUNCTIONAL ESTIMATES OF IMPORTANCE OF FACIAL FEATURES AND OCULOMOTOR BEHAVIOR: A STRAIGHTFORWARD OR AN INTRICATE RELATION?

Armando M. Oliveira[1], Ana D. Silva[1], Isidoro, M.[1], Ricardo G. Viegas[1], Nuno Sá Teixeira[1]
[1]Institute of Cognitive Psychology - University of Coimbra, Portugal
l.dinis@fpce.uc.pt

Abstract

The relative importance of internal features of the face has been a recurrent topic of debate. Two experiments were performed with upright and inverted schematic faces obtained through factorial combination of three internal features. Participants answered on a bipolar graphic scale of expressed hostility -friendliness, while their oculomotor behavior was simultaneously recorded. An averaging model was established in both experiments, which allowed deriving estimates of importance for each feature. Patterns of importance were different for upright and inverse faces. Oculomotor behavior was also different in both situations. However, no simple correspondence between estimated importance and oculomotor indices was observed. Outcomes suggest some trend for an inverse relation between importance and number/ duration of fixations, but also show that this sole trend cannot account for the whole of data.

Attempts at assessing the relative importance of internal features of the face through patterns of eye movements have been made in the past. However, whether indices of oculomotor behavior do reflect the importance of looked at features is mainly the subject of reasonable assumptions (e.g., larger number of received fixations = more importance). Improving on this requires pitting them against legitimate psychological measures of importance. Information integration theory (IIT: Anderson, 1981; 1982) can provide such measures based on the averaging model of integration, which rests on a two parameter representation of the stimulus: its scale value (magnitude), and its importance (weight).

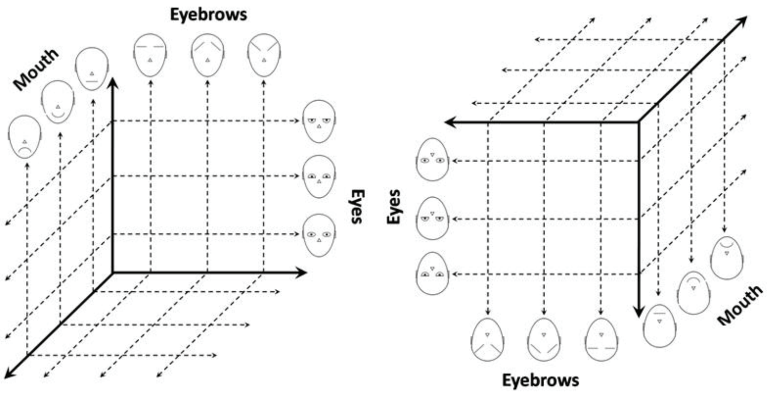


Figure 1 – General diagram of the main design, embodied by the factorial crossing of three schematic facial features (eyebrow, eyes, mouth) within an invariable oval shaped face contour. Points in the intersection of lines correspond to faces (27). *Left:* Upright faces. *Right:* Inverted faces

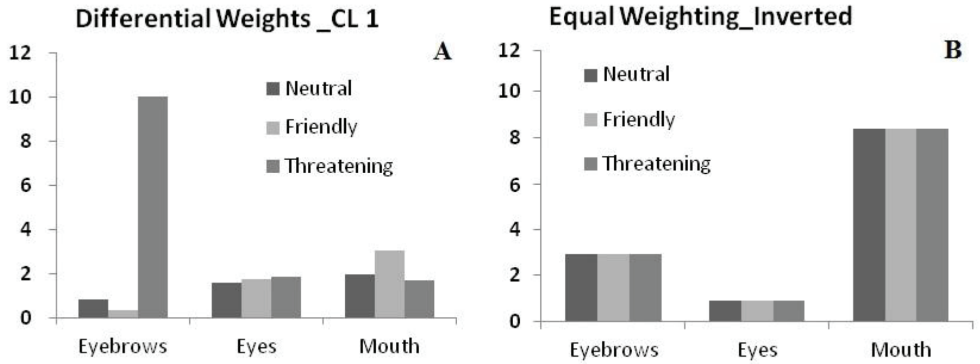


Figure 2 – Weight estimates performed with the R-Average program. Left: Differential-weighting model (14 out of 22 subjects in the up-right condition). Right: Equal-weighting model (all 12 subjects in the upside-down condition).

The data here presented come from two studies on the integration of features of schematic faces whose results have only been partially reported before (Oliveira et al., 2008; 2009). In those studies, an averaging model for the combination of 3 levels of an *eyebrows* factor × 3 levels of an *eyes* factor × 3 levels of a *mouth* factor was found both in upright faces and upside-down faces (see Figure 1), however with differential weighting in the former case (for a major cluster of participants) and equal weighing (i.e., same weight across the levels in each factor) in the latter case. Estimations of importance were extracted with the R-Average program (Masin & Vicentini, 2007), which are shown in Figure 2 (estimations in the up-right faces condition are only presented for the differential-weighting cluster of participants).

While subjects performed on the task of evaluating each face on a bipolar graphic scale of hostility (left pole)-friendliness (right pole), their eye movements and fixations were recorded with an eye-tracker device (Arrington Research). Regions of Interest (RI) in the screen were defined according to the places where the mouth, eyes and eyebrows of the schematic faces would be displayed, and the number and duration of fixations occurring within these regions was averaged across participants. These aggregated indices are presented in Figure 4, with the 27 faces arising out of the factorial design in the abscissa (see Figure 3 for the correspondence between a face number and a given combination of components).

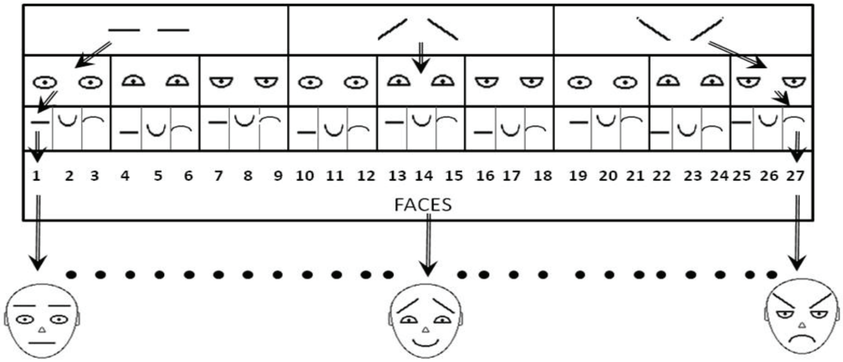


Figure 3. Guidelines for the reading of all graphs presented below: specific combinations of features (eyebrows, eyes, mouth) that each face number (1 to 27) stands for.

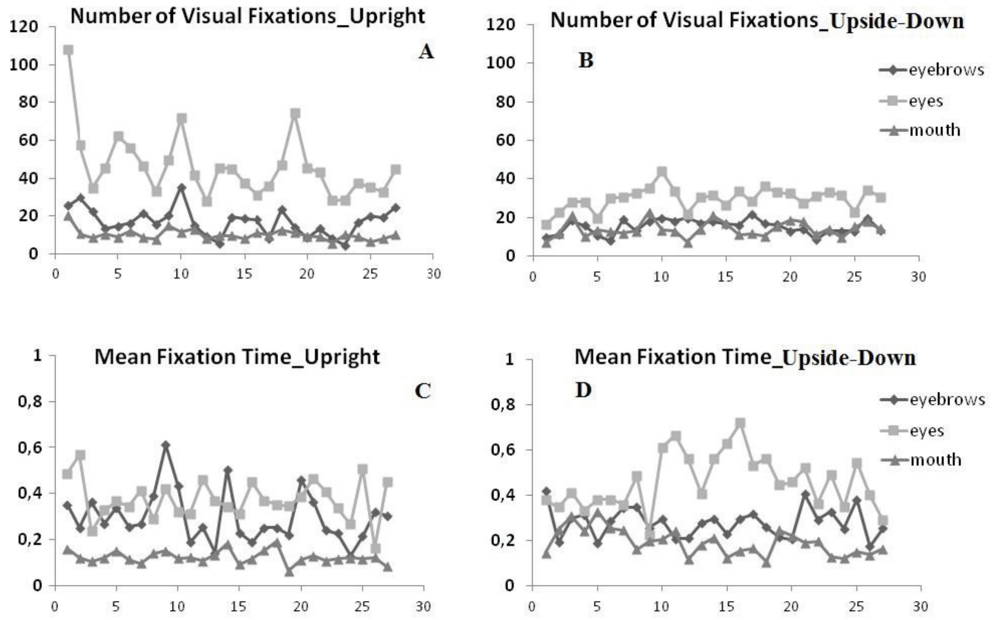


Figure 4. Top Row: Mean number of fixations (ordinate) per feature (curve parameter) in each face (abscissa). Bottom Row: Mean fixation duration in seconds (ordinate) per feature (curve parameter) in each face (abscissa). Left Column: "faces up-right" condition (differential-weighting averaging). Right Column: "faces upside-down" condition (equal-weighting averaging).

Discussion

Just as the integration model for the ratings, oculomotor behavior differs between the upright and upside conditions. Eyes get the larger number of fixations in both conditions, but this number is much larger (and more distinctly above that for the other facial features) in upright than in inverted faces (cf., Figure 4 A and B). A trend for a larger number of fixations given to eyebrows than to mouths can also be seen in the upright faces (where the two lines overlap), but not in the inverted ones. As for the mean fixation durations, they are higher overall in the upside-down than in the upright condition. The major reason for this difference in durations between conditions is the "eyes", which become the target of rather long fixations in inverted faces when "affective" levels of eyebrows (V- or inverse V-shaped) are involved. As a whole, these results indicate more and shorter fixations in upright faces, suggesting that the ocular exploration of upright faces is quicker and finer-grained (which may be the basis for the more complex differential-weighting averaging model).

No simple, straightforward match between the features relative importance as estimated by Functional Measurement (Anderson, 1981; 1982) and the oculomotor indices was observed. Eyes were found to be, on the whole, the less important feature in both conditions (see Figure 2, A and B), while they get the largest number of fixations and the longer fixation durations in both conditions. The distinctive importance of the "threatening" level of eyebrows in the upright condition, which occurs in faces 19 to 27, is not reflected in any way in the plots for either number of fixations or fixation durations (see Figure 4, A and C). Also, the large difference in importance between mouth an eyebrows in the upside-down condition is not obviously marked by any of the oculomotor indices (except for a very slight tendency for longer fixation durations to eyebrows).

Overall, a tentative inverse relation between importance and the derived oculomotor indices is all that can be suggested, whereby the less important factors (eyes) receive more and longer fixations, and the more important factors (such as mouth in the inverted faces) tend to receive less and shorter fixations (not the case in upright faces, though). This might be interpreted (conjecturally) as a link between functional importance and more efficient information processing (non-attentive, pre-attentive?). However, the inverse relation account cannot fit in some of the noticed data (e.g., the absence of an oculomotor effect of the "threatening" level of eyebrows in the upright condition). Given the proposal that different processing mechanisms are called upon by upright and inverted faces (holistic/configural and featural), it cannot be excluded that patterns of relation between psychological importance and oculomotor indices might vary with the kind of facial processing at work.

REFERENCES

Anderson, N. H. (1981). *Foundations of information integration theory*. New York: Academic Press.

Anderson, N. H. (1982). *Methods of information integration theory*. New York: Academic Press.

Lundqvist, D., Esteves, F., & Öhman, A. (1999). The face of wrath: Critical features for conveying facial threat. *Cognition and Emotion, 13(6)*, 691-711.

Lundqvist, D., Esteves, F., & Öhman, A. (2004). The face of wrath: The role of features for conveying facial threat. *Cognition and Emotion, 18(2)*, 161-182.

Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. *Journal of Personality and Social Psychology, 80(3)*, 381-396.

Oliveira, A. M., Teixeira, N. S., Oliveira, M. P., Breda, M. S., Viegas, R., Simões, M. F.& Fonseca, I. B. (2008). Subjective metrics of hostile and friendly facial expressions: an issue with schematic faces. In Bruce Schneider and Boaz Ben-David (Eds.). *Fechner Day 2008. Proceedings of the 24th Annual Meeting of the International Society for Psychophysics* (pp. 117-122). Toronto, CA: The International Society for Psychophysics.

Oliveira, A., Viegas, R., Oliveira, M., Fonseca, I., Jordão, M., Carvalho, C., Mendes, M., Simões, F., Santos, E. (2009). Imputations missing information to incomplete facial expressions: a study with functional measurement. In M. Elliott, S. Antonijecic, S. Berthaud, P. Mulcathy, Bargary, C. Martin and M. Schmidt (Eds.). *Fechner Day 2009: Proceedings of the 25th Annual Meeting of the International Society for Psychophysics* (pp. 353-358). Galway, Ireland: The International Society for Psychophysics.

Vidotto, G., & Vicentini, M. (2007). A general method for parameter estimation of averaging models. Teorie & Modelli, *12(1-2)*, 211-221.

Zalinski, J. & Anderson, M. (1987). *AVERAGE Program & Manual*. San Diego: University of California.

Acknowledgment

This work was supported by Grant PTDC/PSI/73406/2006 from the Portuguese Foundation for Science and Technology

WHAT MAKES MONA LISA SMILE? INVESTIGATING A HALO MODEL OF CONFIGURAL FACE EFFECTS

Armando M. Oliveira[1], Ana Duarte Silva[1], Nuno Sá Teixeira[1], Ricardo G. Viegas[1]
[1]Institute of Cognitive Psychology_University of Coimbra, Portugal
l.dinis@fpce.uc.pt

Abstract

Da Vinci's Mona Lisa is often celebrated for her enigmatic expression, somewhere in between happy and sad. Changes in Mona Lisa's mouth (by superimposing noise on it) have been shown to affect separate ratings of the eyes regarding their happy or sad character, despite these being physically unaltered. Like many other configural face effects, this one has been taken at face value as illustrating a mouth-eyes interaction. The present study investigates an alternative explanation based on the weighted averaging halo model of IIT (Anderson, 1981), which assumes no interaction among components.

The issue of context effects has always been present in the study of faces. Typically, context information makes reference to context outside the face, acknowledging the fact that faces seldom stand in isolation but are commonly met instead in rich interactive settings. The ubiquity of outer contexts also draws attention to the fact that, however important as a communicative device, the face is one of manifold channels of communicative signs (e.g., postures, gestures, prosody, proxemics, verbal utterances, etc.). By far, the dominant question posed by research has concerned the relative importance of contextual information (regarding that of face) for the judgment of emotions (Goodnough & Tinker, 1931; Frijda, 1969; Ekman, Friesen & Ellsworth, 1982; Nakamura, Buck, Kennedy, 1990; for reviews see Wallbott, 1988; Fernandez-Dols & Carroll, 1997). This is also true of the multichannel perspective. Establishing the relative importance of the face channel vis-à-vis other channels, such as voice quality or intonation, has been a standard topic of research along this line (Meharabian & Ferris, 1967; Ekman et al., 1980; Hess, Kappas, & Sherer, 1988). One problem faced by this stream of inquiry, which mostly rests on correlation and regression methods, is lack of valid means for measuring importance (see Anderson, 1982, pp. 262-277; 2001, pp. 275-279, 557-559). Virtually all conclusions drawn over the relative influence of context are thus unwarranted (Anderson, 1989, 165-167). The forthcoming section will illustrate instances of this problem and the solution provided by FM, owing to the weight-value distinction afforded by the averaging model.

However, a more fundamental problem than the one of measurement is whether context and facial expression can be meaningfully addressed as separate informers. Attempts at measuring their respective importance implicitly assume that they do, but this assumption has been questioned on the ground that face and context actually change each other's meaning when in combination (Fernandez-Dols, & Carrol, 1997). The main argument for that rests on phenomena of «vulnerability to reinterpretation» (Fernandez-Dols & Carrol, 1997) whereby changes in context (face) lead to a reconsideration of the meaning attributed to the face (context). Such effects (of which the often cited Kuleshov effect in film editing is an example) actually do little more than comfort the phenomenal impression that contextual information changes the meaning of a face, with no bearing upon the mechanisms whereby context exerts its effect. The interactive interpretation of the face-context relation thus remains no less an assumption than the non interactive, independent account.

Contrary to the idea that this is a strictly empirical matter, distinguishing between the two interpretations cannot be done without adequate model analytic capabilities. This is a documented fact within the FM program, which very early on set forth an alternative to the meaning-change interpretation of context effects (Anderson, 1981, pp. 161-169). This alternative tack involves a two-step integration: all informers (e.g., face and context) are first combined into an overall impression *I*; *I* is then integrated with the particular informer under evaluation (e.g., the face). At the end, the face has undergone a change in phenomenal value, but no intrinsic interaction has occurred between face and context. In a further specification, the integration between *I* and the judged informer was admitted to obey a weighted averaging rule:

$$s' = ws + (1 - w)\,, \tag{1}$$

with *s'* the rating of the in-context informer, *s* its free-context value, and *w* its relative weight in the integration. This formulation corresponds to the 'averaging halo model' of IIT, so called because it rests on the influence of an overall impression over the evaluation of a particular component (Anderson, 1981, pp. 235-244; 1996, 112-115). One noteworthy point is that, as all IIT/FM models, the halo model only assumes the meaning invariance of informers relative to a task-dependent goal and to a judgment dimension, not as a general property of the stimulus. So, «vulnerability to reinterpretation» in the sense argued by Fernandez-Dols ultimately agrees with the FM view that no fixed psychological value preexists in the stimulus. In the meanwhile, it entails no consequence as to how context and the face integrate to produce a judgment along a given response dimension, which is the issue of the meaning-change versus the halo interpretation debate.

Components in a face may be regarded as the surrounding context for a particular component being judged on some dimension. The present study is concerned with such context effects inside, and not outside, the face. It started off from work by Kontsevich and Tyler (2004) over the Mona Lisa's smile. Using the *sfumato* technique, Da Vinci managed to produce an elusive smile which is best seen when not looked at directly (see Livingston, 2000) and tinges the face with an enigmatic expression, somewhere in between happy and sad (Figure 1, leftmost image). By overlaying noise on Mona Lisa's mouth (which produced a change in the perceived smile towards either the happy or the sad pole) Kontsevich and Tyler documented consistent effects upon separate ratings of the eyes on a happy-sad dimension, despite these being physically unchanged (see middle and rightward images in Figure 1, panel A).

Figure 1. *A.* Mona Lisa's smile configural effect: component eyes (unchanged) are perceived more on the happy pole (middle face) or on the sad pole (rightward face) as a function of mouth changes. The leftward face is a gray scale facsimile of the original Mona Lisa's face. *B.* Reproduction of the same configural effect in a synthetic face.

Like many other configural face effects, this one was taken as illustrating a mouth-eyes interaction. However, an alternative explanation exists through the averaging halo model, which assumes no interaction among components. This was investigated in two experiments with synthetic faces (see Figure 1, panel B). In one of them, eyes/brows-related and mouth-related AUs varying along a happy-sad continuum (five levels each) were factorially combined to produce the stimuli-faces. Besides ratings of whole facial expressions on a bipolar happy-sad graphic scale, mouth and eyes components were also separately rated. Thus, when eyes were being evaluated, mouth acted as a context for the evaluation, and vice-versa. The other experiment was similar, except that AUs varied along an anger-happy continuum. Three conditions were additionally created within each experiment, concerning the way faces were presented: up-right, inverted, and with misaligned top and bottom halves. The two later manipulations were expected to increasingly reduce the magnitude of contextual effects, for reasons given below.

Two sorts of predictions from the averaging halo model were tested in the data. The first one concerns the linear relation between the ratings for the eyes or mouth component and the ratings for the overall expression (*I*), as expressed in the equations

$$eyes' = w_e eyes + (1 - w_e)I;\ mouth' = w_m mouth + (1 - w_m)I, \tag{2}$$

with *eyes'* and *mouth'* the in-context judgments, and w_e and w_m the weights of the free-context *eyes* and *mouth*, respectively. One facet of this linear relation is that overall shape of the patterns for the whole expressions should be reflected in the ratings of components. In case the impression *I* itself arises from a linear integration rule, parallelism should then result in the components judgments (Anderson, 2001, 236; Anderson, 1996, 114). However, in case *I* presents systematic deviations from parallelism, those same deviation trends should be mirrored in the contextual ratings.

Results presented in Figure 2 (contrasting judgments of the overall expression, *I*, and of the eyes/brows across presentation conditions) overall agree with this prediction, as revealed by vertical comparisons within each column.

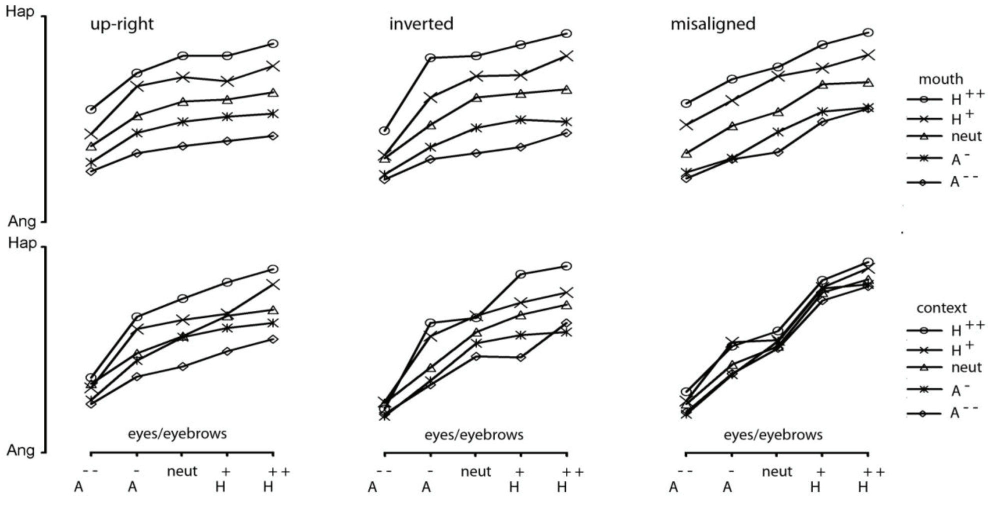


Figure 2. Factorial plots for judgments of the whole expression (top row) and of the eyes/brows component (bottom row) in the Anger-Happy experiment. Mean ratings are on the ordinate, corresponding to a bipolar anger-happy scale (most anger at the bottom, most happiness at the top). Columns correspond to the three presentation conditions.

The extremity weighting effect observed for the leftmost plots in the top row (downward convergence of lines, associated with the 'negative' anger levels) is well replicated in the bottom row. Also, except for specific points, the pattern of inner spacing between lines shows evidence of rough proportionality between judgments of overall expressions and of eyes-in-context. The misaligned condition has the particular property that it exhibits parallelism (sign of a linear integration) for the general expression. According to predictions, parallelism should also emerge in the component judgments, which it does (despite a dramatic reduction of the context effect in this condition, it still reached statistical significance).

The second prediction concerns the effects of presentation mode. The inner relatedness of facial components is widely believed to decrease with face inversion (Yin, 1969) and still more with misalignment of top and bottom face halves (Young, Hellawell, & Hay, 1987). This can be expected to impact on the relative weight of the overall impression (1-*w*), while the context variable affects *I* itself (see Anderson, 1981, 244).

From the (1-*w*) × *I* term in the halo equation it can thus be predicted that plotting context against presentation mode for each level of the rated component will result in graphic linear fans, expressing this multiplicative relation (Anderson, 1981). Plots in Figure 3 strongly support this prediction. Not only do context effects decrease from the upright to the inverted to the misaligned condition (in the expected order, thus), overall they comply with the line fanning typical of multiplying models. Being the less affected by context, the misaligned condition (dashed line) can be seen to operate as a baseline for the fanning. The inverted condition always displays a lesser slope than the up-right condition, concurrent with a decreased relative weight (1-*w*) of the overall impression. These results agree with those obtained by Takahashi (1971; see p. 172) in the domain of personality impression formation. Even if he interpreted them as being against the averaging halo model, they can actually be predicted from it, as pointed out by Anderson (1981, 244). One entailed consequence for holistic processing is that striking examples of holistic effects can actually dispense with configural interaction (essential configurality) and be accounted for in terms of algebraic, non interactive configurality.

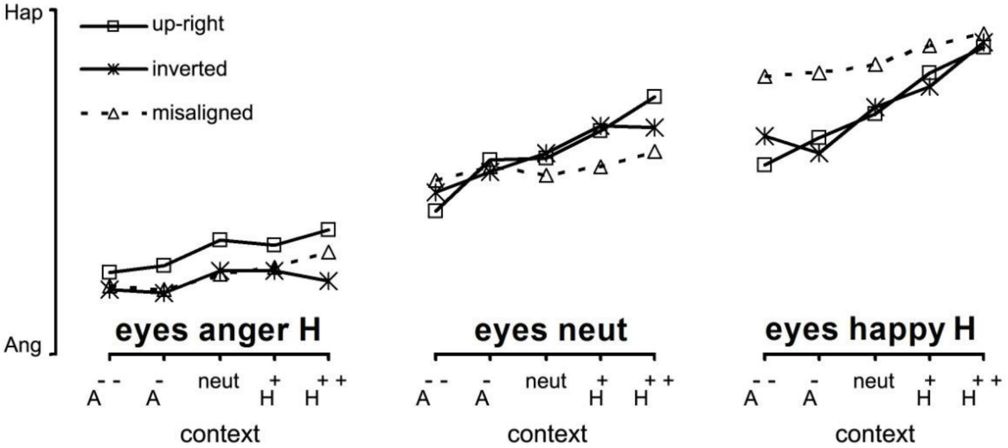


Figure 3. Factorial plots for context (abscissa) × presentation mode (curve parameter). Each graph corresponds to a particular level of the eyes component being targeted for judgment (outcomes presented for only three of the five levels (mean ratings on the ordinate). The context factor is represented by the five levels of the mouth factor, ranging from intense Anger (A++) to intense Happiness (H++).

References

Anderson, N. H. (1981). *Foundations of information integration theory*. New York: Academic Press.

Anderson, N. H. (1982). *Methods of information integration theory*. New York: Academic Press.

Anderson, N.H. (1989). Information integration approach to emotions and their measurement. In R. Plutchik and H. Kellerman (Eds.) *Emotion: Theory, research, and experience. Volume 4: The measurement of emotions* (pp. 133-186). San Diego, CA: Academic press.

Anderson, N. H. (1996). *A functional theory of cognition*. Hillsdale, NJ: Erlbaum.

Anderson, N. H. (2001). *Empirical Direction in Design and Analysis*. Hillsdale, NJ: Erlbaum.

Ekman, P., Friesen, W. V., O'Sullivan, M., & Scherer, K. (1980). *Relative Importance of Face- Body- And Speech In Judgment_of personality and affect.* Journal of Personality and Social Psychology. 1980, 38, 270-277.

Ekman, P., Friesen, W., & Ellsworth, P. (1982). What emotion categories or dimensions can observers judge from facial behavior? In P. Ekman (Ed.), *Emotion in the human face* (pp. 39-55). New York: Cambridge University Press.

Fernandez-Dols, J. M., & Carroll, J. M. (1997). What does a facial expression mean? In J. A. Russell & J. M. Fernandez-Dols (Eds.), *The psychology of facial expression* (pp. 275-294). Cambridge, UK: Cambridge University Press.

Frijda, N. H. (1969). Recognition of emotion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 4, pp. 167–224). New York: Academic Press.

Goodenough, F. L., & Tinker, M. A. (1931). The relative potency of facial expression and verbal description of stimulus in the judgment of emotion. *Comparative Psychology*, *12*, 365-370.

Hess, U., Kappas, A., & Scherer, K. R. (1988). Multichannel communication of emotion: synthetic signal production. In K. R. Scherer (Ed.), *Facets of emotion: Recent research* (pp. 161-182). Hillsdale: NJ: Erlbaum.

Mehrabian, A., & Ferris, S. R. (1967). Inference of Attitudes from Nonverbal Communication in Two Channels. *Journal of Consulting Psychology, 31(3)*, 248–252.

Nakamura, M., Buck, R., & Kenny, D. A. (1990). Relative contributions of expressive behavior and contextual information to the judgment of the emotional state of another. *Journal of Personality and Social Psychology, 59(5)*, 1032-1039.

Takahashi (1971). Effect of the context upon personality-impression formation. *The Japanese Journal of Psychology, 41*(6), 307-313.

Wallbott, H. G. (1988). In and out of context: influences of facial expression and context information on emotion attributions. *British Journal of Social Psychology, 27*, 357–369.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*, 141-145.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *16*, 747-59.

Acknowledgments

This work was supported by Grant PTDC/PSI/73406/2006, from the Portuguese Foundation For Science and Technology

THE OBSERVABLE *R* AND THE UNOBSERVABLE *r:* BRAIN AND RATING RESPONSES IN AN INTEGRATION TASK WITH PAIRS OF EMOTIONAL FACES

Telmo Pereira[1,2], Armando Oliveira[1], Isabel B. Fonseca[3]
[1] *Institute of Cognitive Psychology, University of Coimbra, Portugal*
[2] *Superior College of Health Technology, Polytechnic Institute of Coimbra, Portugal*
[3] *Faculty of Psychology and Educational Sciences, Lisbon, Portugal*
telmo@estescoimbra.pt

Abstract

Participants in this study judged on a graphic rating scale the joint affective intensity conveyed by two emotional faces selectively presented to a single brain hemisphere via a Divided Visual Field technique (DVF). While they did that, EEG was recorded from 6 scalp locations. The emotions considered were joy, fear and anger, varied along three levels of intensity, and each pair of faces might express the same or two distinct emotions. The patterns of integration of the two sources of information were examined both at the level of the ratings and of the brain response (event-related-α-desynchronization:ERD) at each EEG lead. Additive and equal-weighting averaging rules were found, respectively, for the ratings of same-emotion and different-emotion pairs. Additive integration was the predominant finding for α-ERD. Outcomes are discussed with a link to the lateralization of emotional processing and the possible relations between the observable R (ratings) and the implicit neural r.

Information Integration theory (Anderson, 1981; 1982) posits a processing chain whereby observable stimuli (S_i) are turned into their subjective counterparts s_i, which give rise by integration to an implicit response *r*, which is mapped onto an observable R [$r \rightarrow$ R]. In this study we assume that one of the bottom line meanings of *r* might be a neural representation, plausibly in the brain. Perception of facial expressions has been heavily implicated in the debate over hemispheric lateralization of affect/emotion, with almost exclusive reliance on selective unilateral stimulation of one hemisphere by one piece of emotional information (e.g., one emotional face). As suggested by N. Anderson (1996), using tasks requiring the combination of several pieces of information, either provided to a single hemisphere (within-hemisphere integration) or separately to each of the two hemispheres (inter-hemisphere integration) should afford a more natural view of cerebral organization than standard lateralization tasks. This suggestion was taken up in the current study, which used an adapted DVF paradigm with presentation durations (1 sec) allowing to perform the integration task (for DVF paradigms with long exposure durations, see Jansari et al., 2000 and Rodway et al., 2003). However, while Anderson's proposal concerned rating responses in adapted DVF settings, we added to that the simultaneous recording of brain responses (EEG), thus opening up the possibility for a comparison between the observable R and the neural, implicit *r*.

Method

Participants

15 volunteer graduate students (10F, 5M; mean age: 21 + 1.6 years), all of them right-handed, with normal or corrected to normal vision, and naïve regarding the topic under study.

Stimuli

Pictures of emotional faces selected from the JACFEE and JACNeuF databases (Matsumoto & Ekman, 1988). The neutral and highest intensity expression of a given emotion by a same individual were taken as the end poles for a digital morphing at equal 33% steps, providing two intermediate intensities for all the emotions considered: fear, joy and anger (Ekman, 1993). The first and second morphs were used to represent low and intermediate intensity levels of expression. Emotions were taken two by two: each pairing of emotions gave way to 9 stimuli embodying the factorial combination of their intensity levels (3 × 3). The entire set of stimuli also included the 3 x 3 combinations arising out of same-emotion pairs of faces, as well as pairs involving one neutral face and one emotional face (one-factor subdesigns).

Design and Procedure

All stimuli were randomly presented in one single block, using unilateral selective stimulation of one hemisphere at a time (both faces presented to the same visual hemi-field). For each pair of emotions, the overall design corresponded to a full factorial 3 (emotion A) × 3 (emotion B) × 2 (visual hemi-field).

Participants judged the overall intensity conveyed by each pair of expressions on a graphic rating scale, while keeping their eyes at the fixation point. They sat in a recliner in a dimly lit room at a distance of 50 cm from a VGA monitor. The visual angles subtended allowed implementing the divided-half-field technique (5º between the fixation point and the inner edge of the faces pair). All aspects of presentation were managed with SuperLab 4.07, which also triggered the recording of EEG data.

EEG assemblage, data collection and analysis

Six EEG leads (locations F3, F4, T3, T4, P3, P4 on the 10-20 IS) were used, all referenced to Cz. Data were collected at a sample rate of 150 Hz with a band-pass filter of 0.1-35 Hz. Waves were edited offline according to the experimental conditions defined by the factorial designs. Each time epoch included a 2 s baseline period and extended for 10 s after stimuli onset.

A spectral analysis was performed via a fast Fourier transform over the first second following stimulus onset, and α-power was estimated (mean value on the α-band 8.0 – 13.0 Hz). Event-related-α-desynchronization (α-ERD) was then calculated for each epoch, as [(stimulus α-power) - (baseline α-power)]/ (baseline α-power), which reflects the percentage of brain activation in each lead location. To ease up the reading and interpretation of plots, ERD is presented as - (ERD), which results in a positive scale.

Results

Data for pairs of faces expressing the same emotion

The factorial plots obtained from ratings of same-emotion pairs are presented in the two leftmost columns of Figure 1, with left visual field (LVF) presentations on the left and right visual field (RVF) presentations on the right. Overall near-parallelism can be seen in the graphs, supported by lack of statistically significant interactions ($p \geq .242$). This signals an adding-type integration irrespective of the stimulated hemisphere – which, assuming preferential and different lateralization for some of these emotions, would in itself indicate the involvement of inter-hemisphere cooperation.

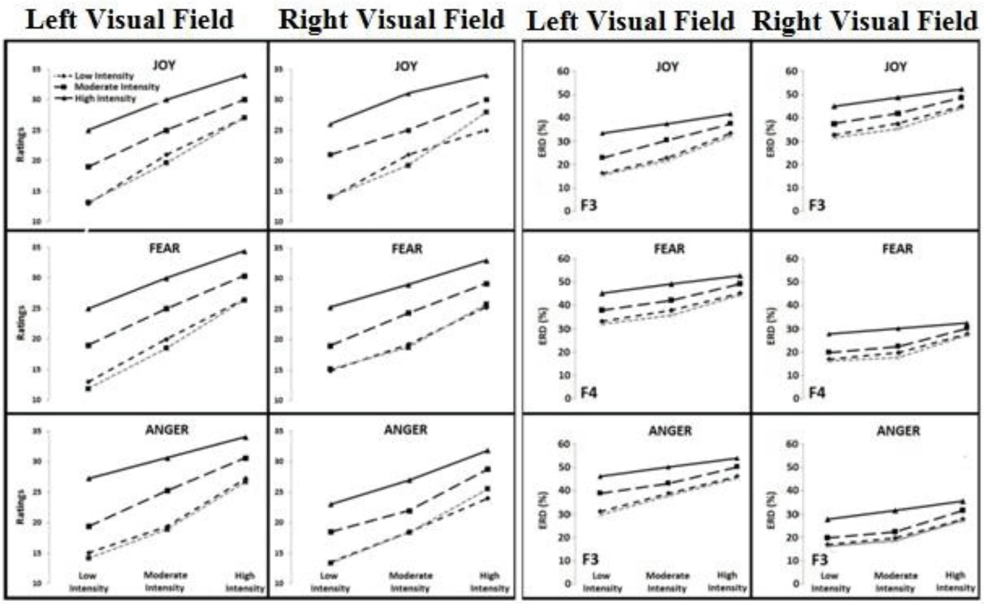


Figure 1. Factorial plots obtained from mean ratings of intensity (left panel) and mean α-ERD at the Frontal Lobe (right panel) for same-emotion pairs (top to bottom: joy, fear, anger). EEG results are given for the lead showing the highest cortical activity for each emotion (Joy and Anger – F3; Fear– F4). The gray line corresponds to the factorial subdesigns (emotional-neutral pairs).

Outcomes for cortical activation are consistent with Davidson's (2004) predictions of a preferential left lateralization of approach-related emotions (higher ERD for Joy and Anger in F3) and a right lateralization of withdrawal-related emotions (higher ERD for Fear in F4) in the Frontal lobe. Concerning the factorial patterns, near parallelism is also prevalent, as in the case for ratings, again supported by non significant interaction terms. Both for ratings and ERD the line corresponding to the subdesigns is also parallel to the lines of the main design, meaning that the intensity of expression of one face is being added to the intensity of the other. An isomorphic adding rule has thus similarly been found at the level of ratings and cortical activation.

Data for pairs of faces expressing different emotions

The factorial patterns obtained from ratings of Joy-Fear pairs (here representing the different-emotion faces pairs) again disclose near-parallelism irrespective of the presentation hemi-field (see Figure 2, top row), well supported by the absence of statistically significant interactions ($p \geq .893$).

However, a clear crossover can now be observed between the lines standing for the one-factor subdesigns, with a much steeper slope, and those corresponding to the main design. The same was observed for all other different-emotion pairs (Fear-Anger and Anger-Joy). This crossover rules out adding and establishes equal-weighting averaging as the integration rule at work in this case (Anderson, 1981, 1982, 1996). As might be expected from an averaging operation, mean ratings in Figure 2 are less extreme and more concentrated around the centre of the response scale than those in Figure 1.

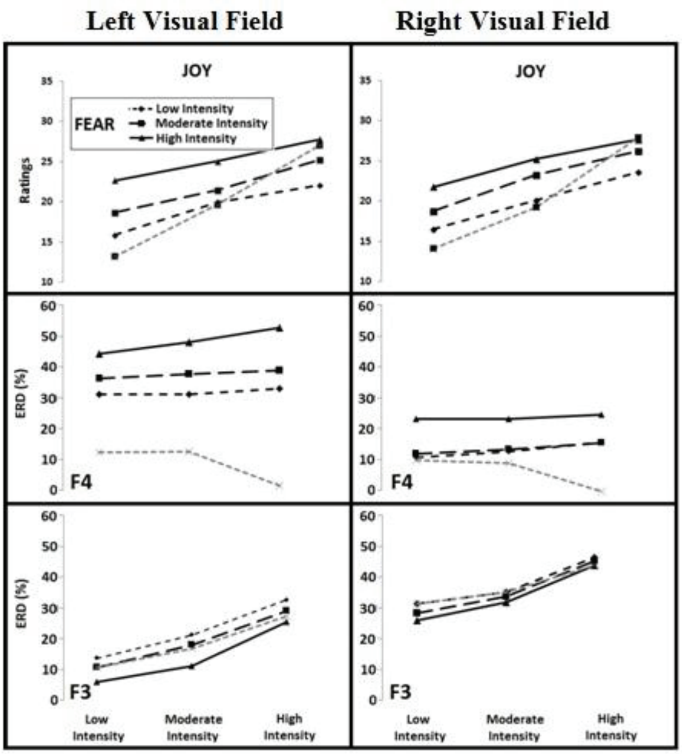


Figure 2. Factorial plots obtained from mean ratings of intensity (upper row) and mean α-ERD at the Frontal Lobe (middle and bottom rows). Data collected for Joy-Fear pairs unilaterally presented to the right or left visual field. In all graphs fear intensity is the curve parameter. Gray lines stand for the factorial subdesigns (emotional-neutral pairs).

As for the patterns of cortical activation involving Joy-Fear pairs, Joy (in the abscissa) presents a rather small effect (though significant) in F4, while the same happens with Fear (which even becomes ns. in the LVF) in F3. A reversal in the functioning of intensities of Fear in F3 is also observed which raises the suggestion of some degree of competition or reciprocal activation between hemispheres (Joy subdesigns, it may be noted, conversely present a diminishing effect in F4). Summing up, each emotion appears to be preferentially treated in a distinct hemisphere, and hemispheres seem to deploy some degree of functional competition among themselves.

Considering both ratings and ERD together, it seems clear that no single hemisphere could support the averaging integration observed in the ratings. Patterns in the ratings and in cortical activation are also not isomorphic, given the different profile evidenced by subdesigns (indicating averaging for the ratings and adding for the ERD). Nevertheless, averaging is an adequate rule to harmonize opposite trends (means are always higher than smaller contributing values and lower than higher contributing values), and could thus well be supported by a trans-hemispheric pattern of collaborative competition (or adversarial collaboration) between the two hemispheres. The major implication for the implicit *r* in this case is that it should be conceived less as a circumscribe neural correlate than as a distributed representation across a bi-hemispheric system.

Discussion

The results presented above concerning cortical activation are partial in two major ways. On the one hand, they do not include the data for temporal and parietal leads, which provided important indications both regarding the dynamics of lateralization and intra-hemispheric organization. On the other hand, they were only presented here for one of the three different-emotion pairs. Nevertheless, despite differences in detail and supplementary conclusions, the missing outcomes do not question the three main provisional conclusions that can be put forward: (1) Additive-type patterns for the combination of emotional information conveyed by faces are prevailing in both ratings and local cortical activation (α-ERD); (2) Evidence of ongoing collaboration, even if sometimes adversarial, between the two hemispheres, is prevailing, and provides the frame for functional lateralization; (3) the implicit r in the [$r \rightarrow$ R] mapping posited by IIT may well sometimes, if not often, correspond to a distributed brain dynamics rather than to a specific neural correlate.

References

Anderson, N. H. (1981). *Foundations of information integration theory*. New York: Academic Press.

Anderson, N. H. (1982). *Methods of information integration theory*. New York: Academic Press.

Anderson, N. H. (1996). *A functional theory of cognition*. Mahwha, NJ: Lawrence Erlbaum Associates.

Davidson, R. J. (2004). What does the prefrontal cortex «do» in affect : perspectives on frontal EEG asymmetry research. *Biological Psychology, 67*, 219-233.

Ekman, P. (1993). Facial Expressions and Emotion. *American Psychologist, 48, 4*, 384-392.

Jansari A., Tranel D., Adolphs R.(2000). A valence-specific lateral bias for discriminating emotional facial expressions in free field. *Cognition and Emotion*, *14*, 341-353.

Matsumoto D & Ekman P (1988). *Japanese and Caucasian facial expressions of emotion (JACFEE)*. San Francisco, CA: San Francisco State University, Intercultural and Emotion Research Laboratory.

Rodway P., Wright L., Hardie S. (2003). The valence-specific laterality effect in free viewing conditions: The influence of sex, handedness, and response bias. *Brain and Cognition, 53,* 452-463

Acknowledgment

Acknowledgments: This work was supported by Grant PTDC/PSI/73406/2006 from the Portuguese Foundation for Science and Technology

ABOUT THE INFLUENCE OF PERCEIVING DISGUSTED, HAPPY AND SAD FACES ON TIME PERCEPTION

Noémie de la Sablonnière, Vincent Laflamme, Katherine Labonté and Simon Grondin
Université Laval, Québec, Canada
noemie.de-la-sablonniere.1@ulaval.ca

Abstract

The aim of the present series of three experiments was to investigate the effect of various emotions (sadness, happiness and disgust) on time perception. In these experiments, temporal bisection tasks were conducted in order to measure the participants' time perception. The various emotions were induced by showing pictures of emotional faces. In Experiment 1, pictures of disgusted and neutral faces were presented and the results revealed an overestimation of time perception when disgusted faces were shown. In Experiments 2 and 3, pictures of happy and neutral faces, and of sad and neutral faces, respectively, were presented. In these experiments, no temporal distortion was found.

The study of the mechanisms behind temporal information processing has a long history. The most popular model used to explain a person's ability to make judgments about time is based on the assumption that there is a "pacemaker-counter" internal clock. It is the accumulation of pulses emitted by the pacemaker that would form the basis for temporal judgments: more pulses accumulated results in longer perceived duration. Judgments about time depend not only on the properties of the internal clock but also on cognitive mechanisms, particularly those tied to attention and memory (Brown, 2008). Indeed, drawing one's attention away from the passage of time results in a reduction of the number of accumulated pulses in the counter and, thus, shortens perceived duration.

Consequently, while it is true that each person has the ability to estimate time more or less precisely, some empirical studies also propose that temporal representations are susceptible to be modulated by external stimuli (Eagleman, 2008). However, one's attention is rarely drawn to the passage of time itself, so it can be difficult to make accurate judgments on it. This is especially true whenever our mind is devoted to more interesting occupations or to emotions.

The effect of emotions on time perception has been gaining more or and more attention from the scientific community. In fact, researchers has been particularly interested in the basic emotions (anger, fear, disgust, happiness, sadness and shame) (Gil & Droit-Volet, 2011a) and the ways to induce them (with emotional pictures or with faces expressing emotions). Even if people cannot know exactly what is felt by the participants in the latter, results showed that the presence and magnitude of the temporal distortions induced depend on which emotion is presented.

Gil and Droit-Volet (2011a) conducted an experiment during which they presented disgusted faces and neutral faces. Their goal was to compare the effect of both types of stimuli on time perception. Their results showed that no temporal bias was being induced by disgusted faces. Gil and Droit-Volet concluded that disgusted faces do not send any socially relevant message. Indeed, according to them, even though disgust is a “high-arousal” and “unpleasant” emotion just like anger and fear, it does not influence time perception in the same way as the two latter emotions because it sends a different social clue to the observer. Since the basic function of disgust is to protect oneself from poisoning, no immediate action would be necessary for a person observing another displaying such an expression (Droit-Volet & Gil, 2009). To the knowledge of the authors, this study has not yet been replicated.

In the study of Droit-Volet, Brunot and Niedenthal (2004), the authors compared the effect induced by happy expressions (corresponding to Duchenne smile (Duchenne, 1990)) with the ones induced by faces expressing no emotion. Because happiness is a “low-arousal” emotion, the observed temporal overestimations were of low magnitude, but still statistically significant. This study has been replicated by other authors. Effron, Niedenthal, Gil and Droit-Volet (2006) presented faces expressing Duchenne smiles and obtained the same results as Droit-Volet et al. However, Gil and Droit-Volet (2011a) and Tipples (2008) studies showed no temporal distortion from the presentation of faces expressing happiness (non-Duchenne smile). Gil and Droit-Volet posited that the absence of significant results was caused by the fact that the images of faces expressing happiness they used in their experiment didn't show a Duchenne smile. Therefore, they weren't interpreted as genuine smiles by the participants. Conversely, when a person shows a genuine smile, it is interpreted by the observer as an invitation to approach and to make contact. This reaction would in turn increase the observer’s arousal and cause him or her to overestimate time (Droit-Volet & Gil, 2009).

Droit-Volet et al. (2004) also explored the influence of sadness on time perception. In order to do this, they presented faces expressing sadness as well as faces expressing no emotion. In that experiment, a marginally significant overestimation of time caused by sad faces was observed. They concluded that perceiving sadness in someone else sends the message to the observer that it is necessary to give assistance to that person. Consequently, hormones would be secreted in the observer’s organism in order to help the sad person. Those results were replicated by Gil and Droit-Volet (2011a). The latter study even found a statistically significant overestimation. The authors nuanced their findings by specifying that the magnitude of the temporal overestimation brought by the presentation of sad faces was greater in five-years-old children than in older subjects. Thus, they hypothesized that, due to the social inhibition brought by sadness, children eventually learn to suppress that emotion. Alternatively, this attempt to suppress the emotion of sadness could also draw attention away from time and, thus, reduce the number of pulses that are accumulated in the counter component of the internal clock and the resulting temporal overestimation.

The general way by which such stimuli might influence a person's time perception usually takes the form of empathy and pro-social sensitivities. Indeed, Gil and Droit-Volet (2011a) posited that the temporal distortions induced by emotional faces had very specific social utility. Therefore, one's ability to feel empathy may modulate the effect of such stimuli on time perception. The interpersonal reactivity index (IRI) is a questionnaire measuring a person's empathic abilities (Davis, 1980) with four different subscales (each measuring a specific component of empathy). The "perspective-taking" subscale measures one's tendency to adopt another person's point of view. The “empathic concern” subscale measures one’s ability to feel others’ emotions and to feel concerned by the other’s distress. The “personal distress” scale evaluates the anxiety level felt by the persons when they give some assistance to a distressed person. The “fantasy” subscale measures one’s tendency to get caught up in fictional story and to imagine oneself in the same situation as a fictional character. Those additional measures sought to link the magnitude of the effect of emotions on time perception to the different specific measures of empathy.

In order to elaborate further on Droit-Volet et al. (2004), Effron et al. (2006), Gil and Droit-Volet (2011a) and Tipples' (2008) hypotheses, it is imperative first to replicate their findings. Replicating findings is a key component of the scientific method, particularly when studying emotions. Indeed, a large interindividual and intraindividual variability can occur. Moreover, Gil and Droit-Volet (2011b) themselves found that the nature of a task can itself heavily influence the overall results (over- vs. underestimation).

During the course of this study, three experiments using a temporal bisection task were conducted, which involved: (1) faces expressing disgust and neutrality, (2) faces expressing happiness and neutrality and (3) faces expressing sadness and neutrality. The present experiments closely followed the procedures of Gil and Droit-Volet (2011a). In addition, the IRI questionnaire was completed by each of the participants.

General Methods

Participants
For each of the three experiments, 16 students or employees of Laval University were recruited. Each of them received 5$ per session (20$ total). Each session lasted approximately 30 minutes. The participants in Experiment 1 were 21 to 37 years old (M = 25.44). Nine of them were women and 7 were men. In Experiment 2, participants were 20 to 38 years old (M = 26.81). Ten of them were women and 6 were men. In Experiment 3, the age of participants (10 women and 6 men) ranged from 20 to 38 years old (M = 23.00). All participants took part in only one experiment.

Material
Each participant performed his or her task individually in an isolated room at the perception laboratory at Laval University. The room was dimly lit by a small desk lamp so that the participants could see the computer screen clearly. The program that presented the stimuli and recorded the participants' responses was designed using the E-prime 2.0 software. In order to respond, the participants had to press either the "1" key or the "3" key of the keyboard. The visual stimuli (5 neutral faces, 5 faces expressing disgust, 5 faces expressing sadness and 5 faces expressing happiness (non-Duchenne smile)) were sampled from the “Montreal Set of Facial Expressions” (Beaupré, 2000). They were presented on 7 X 10 inches computer display (IBM Think vision with an 80 Hz refresh rate). The experiment was under control of an IBM Pentium 4 computer.

Procedure
At the start of the first session, each participant was invited to complete the IRI questionnaire. Once they were done, the participants had their task explained to them and they were given an instruction sheet. They were then led to the room where they performed their task. Just like in Gil and Droit-Volet's studies (2011a), a temporal bisection task was used. In each experiment, there were 4 identical sessions lasting approximately 30 minutes. At the start of each session, the participants completed a learning block during which each of the two anchor durations (400 and 1600 ms) was presented 10 times to them. Each anchor duration was delimited by the onset and the offset of a picture of a neutral female face. After each presentation of the anchor durations, there was a 3500-ms delay. After the learning block, the participants completed a 14-trial practice block and were given feedback on their performance after each trial. The participants then had to complete 4 experimental that each were comprised of 70 trials. During each trial of the practice and the experimental blocks, the participants were presented a comparison interval which would last either 400, 600, 800, 1000, 1200, 1400 or 1600 ms. These intervals were delimited by the onset and offset of a visual stimuli which could either be a picture of a neutral face or a picture of an emotional face. For each type of visual stimuli, there were 3 male and 2 female faces. Immediately after the offset of the comparison interval, the participants had to respond either "short" by pressing "1" if they thought the comparison interval was more similar to the shorter anchor duration or "long" by pressing "3" if they thought it was more similar to the longer anchor duration.

Results

The proportion of "long" responses was calculated for both conditions in each of the three experiments (neutral faces and emotional faces). Nonlinear regressions using the cumulative Gaussian function were then conducted on the proportion of "long" responses. From these regressions, two parameters were extracted: the mean and the standard deviation parameters of the regression curve. The mean parameter was used as an estimator of the point of subjective equality (PSE) and the standard deviation parameter divided by 1000 ms was used as an estimator of the Weber Ratio (WR). The larger the PSE, the shorter time is perceived. A large value of WR is associated with a lesser sensitivity to time. For each experiment, a paired samples t-test (emotional vs. neutral) was conducted on the PSE and on the WR. Figure 1 summarizes the mean PSE obtained in each condition of each of the three experiments.

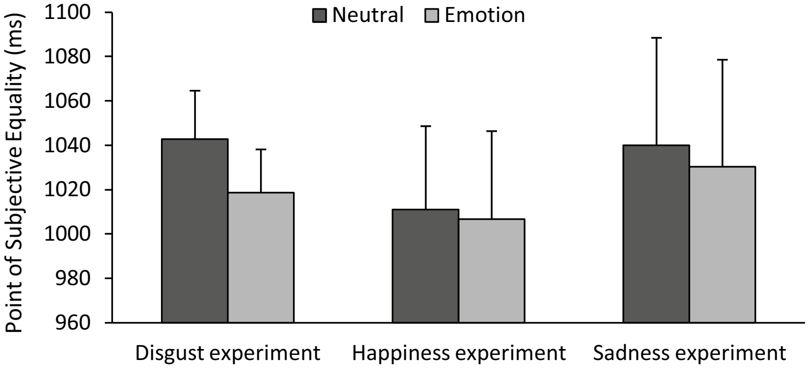


Figure 1. Mean point of subjective equality for each condition of each experiment. Bars are standard error.

Experiment 1: Disgust

The PSE of the neutral faces condition was significantly higher than that of the disgusted faces condition, $t(15) = 3.326$, $p = 0.005$, $d = 0.877$. For the WR, the difference was not significant ($p = 0.298$).

Experiment 2: Happiness

The difference between the neutral faces and happy faces conditions was not significant. This findings applies to both PSE ($p = 0.603$) and WR ($p = 0.680$).

Experiment 3: Sadness

The difference between the neutral faces and sad faces conditions was not significant. This findings applies to both PSE ($p = 0.342$) and WR ($p = 0.485$).

IRI scores

To test the hypothesis that an individual's empathic abilities modulates the effect of the presentation of disgusted faces on his or her time perception, correlation indexes (Pearson's r) were calculated between the difference of PES between the disgust and neutral conditions and the score at each of the IRI subscales. None of the correlation indexes were found to be significantly different from 0.

Discussion

Experiment 1: Disgust

The presentation of disgusted faces induces, at least in the current study, an overestimation of time as compared to the presentation of a neutral faces. This could mean that this emotion, when expressed by another person, results in an increase of the observer's arousal. This would seem to contradict Gil and Droit-Volet's (2011a) findings where the very same stimuli did not induce any temporal distortion and, therefore, did not influence the observer's internal clock. They concluded that disgusted faces do not send any socially relevant message.

Given that the results of the present experiment are inconsistent with Gil and Droit-Volet's

(2011a) findings about the effect of disgust, perhaps the expression of disgust in another person is interpreted by an observer as a social clue. While the expression of disgust is often associated with the sight of rotten food or other types of contaminants, it can also arise when a person witnesses another committing immoral acts (Chapman, Kim, Susskind, & Anderson, 2009). The latter is known as moral disgust. The participants of this experiment could very well have interpreted the disgusted faces as a sign of disapproval. Consequently, their time perception would have been influenced in a manner that is similar to that of angry faces (Gil & Droit-Volet, 2011a,b): their level of arousal would have been increased as they prepared for eventually defending themselves.

Experiment 2: Happiness

Given the present experiment's results, it would seem that presenting faces expressing happiness does not induce any statistically significant temporal bias in the observer. This finding supports Droit-Volet and Gil (2009) and Gil and Droit-Volet's (2011a) hypothesis. Indeed, those researchers thought that the presence or the absence of temporal distortion was explained by the presence or the absence of Duchenne smile, respectively. This explanation could also hold for the present experiment since, like in Gil and Droit-Volet, the images were taken from the "Montreal Set of Facial Displays of Emotion" (MSFDE). Thus, the way by which people interpret happy expressions would depend on the type of smile expressed. Since people learn to smile in many other situations than when they genuinely feel happy, it is possible that they learn to distinguish true expressions of happiness from "social" smiles. These two expressions would then have two different significations.

Experiment 3: Sadness

According to the results from the present experiment, there is no distortion in time perception when participants are presented pictures of sad faces, as compared to neutral faces. This finding contradicts those of Droit-Volet et al. (2004) and of Gil and Droit-Volet (2011a). However, just as Gil and Droit-Volet proposed, it could be that when the subjects are presented sad faces, their attention is drawn away from time in order to repress their negative emotion. This would explain why some studies find temporal distortions of very low yet statistically significant magnitude while others find no significant effect of the sad faces on time perception.

IRI scores

The empathic abilities of the participants were measured by the various subscales of the IRI. Consequently, it was expected that there would be a relationship between those scores and the magnitude of the temporal biases measured in the present experiment. This was not the case, since there were no statistically significant correlations between the scores at the various subscales and the differences in PSE. These results would therefore suggest that empathy have no, or perhaps very limited, influence on how emotional faces cause distortions in a participant's time perception. Then again, given that none of subjects lacked empathy in a pathological way, it could still be possible that the link between empathy and the effect of emotional faces on time perception is a dichotomic one and not a matter of degree. In other words, one would only need a minimal amount of empathic abilities to have his or her time perception influenced by other people's facial expressions. Therefore, it would be interesting to conduct a similar experiment and to recruit participants who suffer from empathic disorders (psychopathy, for example).

Another explanation for this lack of statistically significant link could be that the durations studied here were too short for the participants to process emotional expressions consciously. It is thus possible that the participant's reaction to the emotional faces was automatic and physiological in nature. Since empathy is processed consciously, it would have little impact on time perception.

WR scores

As was expected, there was no statistically significant effect of emotional faces on the WR. These results support the idea that these types of stimuli do not influence a person's sensitivity to time. In other words, the precision of one's temporal judgments is not affected by the emotional stimuli.

Final comments

Even though the main goal of the present study was to replicate Gil and Droit-Volet's (2011a) findings, some methodological differences were nevertheless introduced. Those differences could have improved or hindered the present study's validity. First of all, the present experiments took a much greater number of observations per point on the psychometric functions (80 vs. 9) in order to increase the robustness of the estimators, the PSE and the WR. Secondly, the images that delimited the anchor durations shown during the learning block depicted neutral faces instead of pink ovals as in Gil and Droit-Volet's studies. Thirdly, each of the anchor durations was presented 10 times successively during the learning of the present study rather than 4 times each in a random order. This change was done so that participants may learn the anchor durations more easily and more accurately. Fourthly, both the stimuli and the participants were of both genders. This aspect might explain why the present study yielded smaller effects than those of Gil and Droit-Volet since the latter used only female participants and female stimuli. Finally, the statistical method employed in the present study consisted in conducting non-linear regressions on the psychometric functions using the cumulative Gaussian function. In contrast, Gil and Droit-Volet's estimated their parameters by linear regression near the inflection point of the psychometric function. The method employed in the present study is more accurate and precise.

References

Beaupré, M.G., Cheung, N., & Hess, U. (2000). *The Montreal Set of Facial Displays of Emotion* [Slides]. (Available via Ursula Hess, Département de psychologie, Université du Québec à Montréal) http://www.psychophysiolab.com/download.php

Brown, S.W. (2008). Time and attention: Review of the literature. In S. Grondin (Ed.). *Psychology of time* (pp.111-138). Bingley, UK: Emerald. doi: 10.1016/B978-0-08046-977-5.00004-1

Chapman, H. A., Kim, D. A., Susskind, J. M., & Anderson, A. K. (2009). In bad taste: Evidence for the oral origins of moral disgust. *Science*, *323*(5918), 1222–6. doi:10.1126/science.1165565

Davis, M.H. (1980). A multidimensional approach to individual differences in empathy. JSAS Catalog of Selected Documents in Psychology, 10, 85.

Droit-Volet, S., Brunot, S., & Niedenthal, P. (2004). Perception of the duration of emotional events. *Cognition & Emotion*, *18*(6), 849–858. doi:10.1080/02699930341000194

Droit-Volet, S., & Gil, S. (2009). The time-emotion paradox. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *364*(1525), 1943–53. doi:10.1098/rstb.2009.0013

Duchenne, G.B.A. (1990). The mechanism of human facial expression. In R.A. Cuthbertson (Ed. And Trans.). Cambridge, UK: Cambridge University Press. (Original work published in 1862).

Eagleman, D. M. (2008). Human time perception and its illusions. *Current Opinion in Neurobiology*, *18*, 131–6. doi:10.1016/j.conb.2008.06.002

Effron, D. A, Niedenthal, P. M., Gil, S., & Droit-Volet, S. (2006). Embodied temporal perception of emotion. *Emotion*, *6*, 1–9. doi:10.1037/1528-3542.6.1.1

Gil, S., & Droit-Volet, S., (2011a). How do emotional facial expressions influence our perception of time? In S. Masmoudi, D. Yan Dai & A. Naceur (Eds). *Attention, Representation, and Human Performance: Integration of Cognition, Emotion and Motivation*. London: Psychology Press, Taylor & Francis.

Gil, S., & Droit-Volet, S. (2011b). "Time flies in the presence of angry faces"… depending on the temporal task used! *Acta Psychologica*, *136*(3), 354–62. doi:10.1016/j.actpsy.2010.12.010

Tipples, J. (2008). Negative emotionality influences the effects of emotion on time perception. *Emotion, 8*, 127–31. doi:10.1037/1528-3542.8.1.127

Acknowledgement. This research was made possible by a research grant from the Social Sciences and Humanities Research Council of Canada awarded to SG.

ABOUT THE EFFECT OF FEAR AND DISGUST ON TIME PERCEPTION

Vincent Laflamme, Noémie de la Sablonnière, Vanessa Fournier, Simon Grondin
Université Laval, Québec, Canada
vincent.laflamme.1@ulaval.ca

Abstract

The goal of the present study was to explore the effects of emotions on time perception. More specifically, the effects of three types of stimuli – faces expressing fear (FEF), disgust (FED) and pictures of disgusting food (PDF) – were compared using a temporal bisection task. The results showed that, compared to FEF, time intervals seem significantly shorter when participants are exposed to PDF. However, no other differences were found between conditions. Additionally, there was no statistically significant difference in Weber ratio (WR) between all of the conditions.

Time perception is an integral part of the processing of information required for our day-to-day life (Chambon, Gil, Niedenthal & Droit-Volet, 2005). Temporal properties of various events have to be encoded in order to build temporal representations which, in turn, can be used to make decisions (Droit-Volet, 2005). One's ability to perceive time accurately, to allocate it to various activities and to remember how much time has passed is crucial to his or her social adaptation.

So far, many models of temporal information processing have been elaborated in order to describe and explain the mechanisms underlying perception and estimation of time (Grondin, 2001). To this day, the most popular model used for explaining how time is perceived or estimated probably remains that of Gibbon, Church and Meck (1984). This model, called the scalar expectancy theory, is based on the postulate that a "pacemaker-counter" type of internal clock exists in our mind. It posits that one's perception of a given time interval depends on the number of pulses accumulated in the counter component: the larger the number of accumulated pulses is, the longer the interval will seem. These pulses are emitted regularly by a pacemaker and have to pass through a component called the "attentional switch". In addition to the properties of the internal clock components, temporal judgments could also be influenced by cognitive processes, especially those associated to attention and to memory.

While it is true that each person has the ability to estimate time accurately, empirical research has shown that temporal representations are susceptible to be altered by many types of stimuli (Eagleman, 2008). In particular, the question as to how stimuli illustrating or generating emotions influence time perception is presently the subject of much research. Indeed, it has been demonstrated several times that when a person feels or recognizes some emotions, his or her perception of time is disturbed (Gil & Droit-Volet, 2011a).

Research suggests that certain emotions can potentially accelerate one's internal clock. This will increase the amount of pulses that are stored in the counter during a time interval, consequently causing an overestimation of time (Gil & Droit-Volet, 2011b). Droit-Volet and Gil (2009) explain the acceleration of the internal clock by the intermediate of an automatic process causing an increase of arousal. This arousal facilitates the anticipation of future events and the preparation to take action. The attentional switch is another structure of the internal clock that can potentially be modulated by emotions (Droit-Volet & Gil), and more particularly by attentional processes. Whenever a person's attention is drawn away from time, some pulses can be lost because of the late closing of the switch at the beginning of the time interval. If less pulses are accumulated, time is underestimated (Gil, Rousset & Droit-Volet, 2009). In that case, attention will again be drawn away from time because of the person's need to adopt a reflective attitude. All those different reactions and temporal distortions are explained by the emotions' signification and by the necessity to react rapidly or not. Thus, the impact of emotion on time perception would depend on their fundamental meaning and their underlying mechanisms (Droit-Volet & Gil).

The goal of the present study was to examine the influence of emotional stimuli. More specifically, the effect of images that induce disgust and fear were of interest. Participants were subjected to three types of stimuli: pictures of disgusting food (PDF), faces expressing disgust (FED) and faces expressing fear (FEF) during a temporal bisection task. It was predicted that an interval would be perceived as shorter when people are exposed to PDF, compared to when they are exposed to FEF. According to our review of the literature, these two conditions represent extremes, or "poles", in term of temporal distortions that can be induced. Previous studies suggest that PDF condition would induce underestimation of time while FEF condition would induce overestimation (Gil & Droit-Volet, 2011a; Gil *et al.,* 2009). Moreover, we also expected that duration would seem longer in FEF than in FED condition. As Gil and Droit-Volet posited, seeing FEF increases one's arousal more than seeing FED, thus inducing an overestimation of time. Finally, we expected that, compared to FED, a time interval would seem shorter when PDF is presented. Indeed, as observed in Gil *et al.*'s study, since PDF influence the attentional processes, that kind of stimuli would induce an underestimation of time. Also, according to Gil and Droit-Volet, the presentation of FED should not induce temporal distortion.

Method

Participants

Eight men and 18 women (mean age = 23 years) were recruited at Laval University. They received 10$ for their participation. In order to be recruited, volunteers had to be 18 to 40 years old, have no uncorrected vision disorders and no neurological or psychological disorder that would require medication. No participant took part in other time perception experiment involving the presentation of emotional pictures.

Material

Each participant performed their task individually in an isolated room at the perception laboratory at Laval University. The room was dimly lit with a small desk lamp so that the participants could see the computer screen clearly. The program that presented the stimuli and recorded the participant's responses was designed using the E-prime 2.0 software. The pictures of faces expressing fear or disgust were taken from the « Montreal Set of Facial Expression » image bank (Beaupré & Hess, 2000). The pictures of disgusting food, which were validated by Rousset, Deiss, Juillard, Schlich and Droit-Volet (2005) were taken from the «SU.VI.MAX» manual (1994). They were also used in Gil *et al.*'s (2009) study.

Procedure

Each participant completed a temporal bisection task and was subjected to six experimental conditions (two sets of intervals and three types of pictures). There were two counterbalanced sessions of about 30 minutes each. During each session, there were two phases: a learning phase and an experimental one. During the learning phase, participants were presented each of two images (anchor durations) 10 times and were instructed to memorize their duration. They were marked by the onset and the offset of a visual stimulus (a blue circle on a black screen). During the experimental phase, the two anchor durations as well as five others were presented. Durations in this phase were marked by the onset and the offset of one of the three emotional stimulus types (PDF, FED or FEF). In one of the two sessions, the short anchor duration lasted 376 ms and the long anchor duration 1000 ms (short base duration, < 1 s). The five other durations were 480, 584, 688, 792 and 896 ms long. In the other session, the anchor durations were 1000-ms and 1624-ms (long base duration, > 1 s). The five other durations were 1104, 1208, 1312, 1416 and 1520 ms long. Immediately after the offset of the comparison interval, the participants had to respond either "short" by pressing "1" if they thought the comparison interval was more similar to the shorter anchor duration or "long" by pressing "3" if they thought it was more similar to the longer anchor duration.

Results

A 7-point psychometric function was first drawn for each experimental condition and for each participant. Then, a non-linear regression using the cumulative Gaussian function was conducted on each psychometric function.

In order to measure the perceived duration during a temporal bisection task, one usually calculates the point of subjective equality. It is defined as the theoretical duration that would obtain exactly 50% of "long" response. The mean parameter extracted from the non-linear regressions was used to estimate the PSE. Furthermore, those PSEs were divided by the average of the two anchor durations so that the short interval set and the long interval set conditions might be compared. The resulting estimator was called the proportional point of subjective equality (PPSE).

As an index of the participants' sensitivity, the Weber ratio (WR) was used. It is defined here as the standard deviation parameter of the non-linear regressions divided by the average of the two anchor durations.

The PPSE and the WR, the dependent variables of this study, were calculated for each participant, for each type of stimuli (PDF, FEF, FEP) and for each set of durations (376 to 1000 ms and 1000 to 1624 ms). The mean PPSE in experimental each condition is illustrated in Figure 1.

The R^2 coefficient was calculated for each regression. The goodness-of-fit values were relatively high; the mean R^2 value was 0.95. A 2 (sets of durations) X 3 (types of stimuli) repeated measures ANOVA with the Greenhouse-Geisser correction ($\varepsilon = 0.751$) was conducted on the PPSE. The ANOVA yielded a statistically significant (at $\alpha = 0.05$) main effect of stimulus type $F(1.5, 37.6) = 4.10$, $p = 0.03$, $\eta_p^2 = 0.14$. The main effect of the set of durations used and the interaction of the two factors were not statistically significant ($p = 0.65$ and $p = 0.34$, respectively).

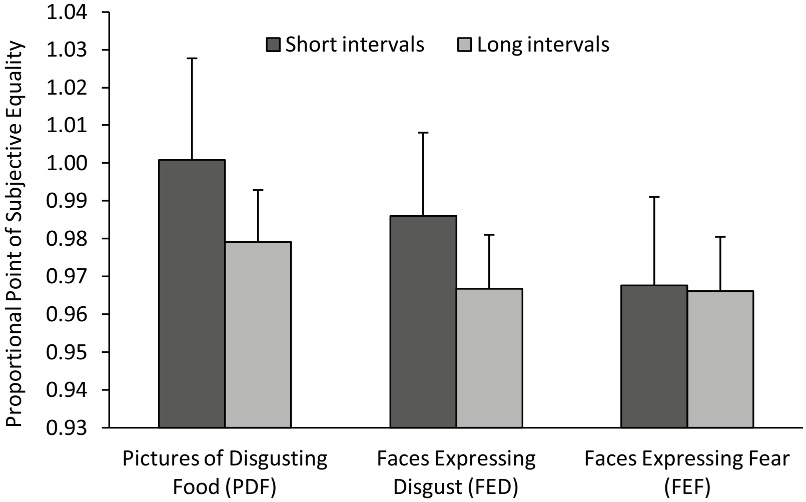


Figure 1. Mean proportional point of subjective equality for each condition of each experiment. Bars are standard error.

Following the ANOVA, three unidirectional paired-sample t-tests were conducted using Bonferroni adjusted alpha levels of 0.0167 per test (0.05/3). The first t-test comparing the PPSE of the FEF (M = 0.967, SD = 0.012) and the PDF (M = 0.990, SD = 0.012) conditions showed a statistically significant difference, $t(25) = 0.023$, $p = 0.04$, $d = 0.39$. The second t-test, which compared the PPSE of the FEF and the FED (M = 0.976, SD = 0.012) conditions, yielded no significant difference, $p = 0.16$. The last t-test, which compared the effect of FED and PDF also yielded no significant difference, $p = 0.19$.

A 2 (sets of durations) X 3 (types of stimuli) repeated measures ANOVA was conducted on the WR. The ANOVA showed no significant effect of the type of stimulus, $p = 0.53$, no significant effect of the set of durations used, $p = 0.73$, and no significant interaction effect, $p = 0.16$. The mean WR in experimental each condition is illustrated in Figure 2.

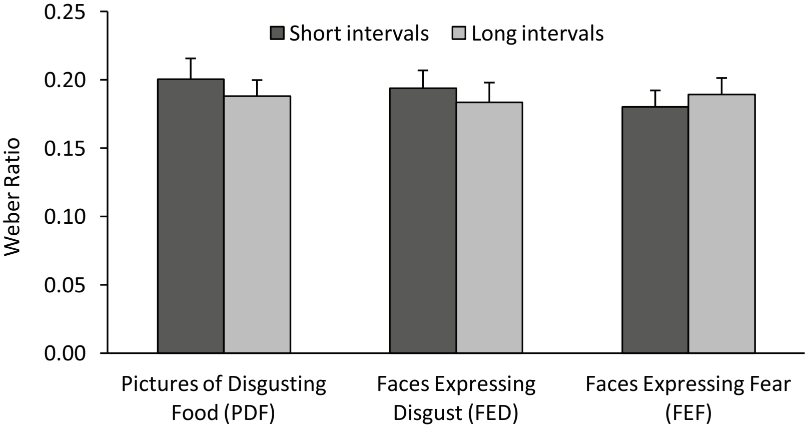


Figure 2. Mean Weber ratio for each condition of each experiment. Bars are standard error.

Discussion

Proportional Point of Subjective Equality

The PPSE difference between the PDF and FEF conditions turned out to be significant. This result suggests that a person perceives time as shorter when he or she is exposed to PDF than when exposed to FEF. These results are consistent with those reported by Gil *et al.* (2009) and by Gil and Droit-Volet (2011a). Indeed, in the 2009 study, the presentation of PDF had induced an underestimation of time. According to the authors, the exposition to PDF draws the participants' attention away from time itself. Additionally, the results from Gil and Droit-Volet indicate that participants who are subjected to FEF tend to overestimate time. The purpose of this alteration of perception would be for the observer to be prepared to react rapidly. In short, the results of the present experiment tend to support the idea that FEF and PDF cause different effects leading to opposite temporal distortions.

Contrary to what was expected, there was no significant PPSE difference between FED and FEF or between FED and PDF. However, according to Gil and Droit-Volet (2011a) and Gil *et al.* (2009), these differences should have been significant. The following three arguments might explain why our results differ from previous data.

First, the experimental designs differed slightly. In Gil and Droit-Volet (2011a) and Gil *et al.* (2009), a greater number of participants were recruited than in the present experiment. This means that the former studies' statistical tests might have been more powerful than those used in the present experiment. However, those studies only involved 9 observations per data point on the individual psychometric functions, compared to 21 in the present experiment. Thus, all other things being equal, the coefficients that were estimated from the psychometric functions should have been more precise and accurate in the present experiment. Since both of the differences in experimental designs yielded opposite effects, it's not clear whether they were the cause of the differences in results.

A second argument that might apply in this case regards the participants' gender as well as the gender of the faces shown. It has been shown in the literature that one's sensitivity to another person's emotions is greater when that other person is of the same gender (Chambon *et al.*, 2005).

The third and last argument is related to the statistical methods employed. As was previously stated, a non-linear regression was used for the present study in order to estimate both of the parameters of interest. Gil and Droit-Volet (2011a) and Gil *et al.* (2009), on the other hand, conducted linear regressions near the inflexion point of the psychometric functions. Since the former method is more accurate and precise than the latter, this could help explain the discrepancies between their results and the present results.

Weber ratio

As for the Weber ratio, there was no statistically significant effect, neither for the type of stimuli (emotions) nor for the duration set (376-1000 ms vs. 1000-1624 ms). These results support the affirmation according to which one's sensitivity to time –the precision of their temporal estimations is not affected by emotional faces. Furthermore, the absence of a significant duration set effect justifies, at least partially, the use of the scalar expectancy theory’s theoretical framework. According this theory, the Weber's law (scalar property) for time should hold, which is consistent with our data (the non-significant effect).

References

Beaupré, M.G., Hess, U., 2000. Cross-cultural emotion recognition among Canadian ethnic groups. *Journal of Cross-Cultural Psychology, 36*, 355–370.

Chambon, M., Gil, S., Niedenthal, P. M., & Droit-Volet, S. (2005). Psychologie sociale et perception du temps : l'estimation temporelle des stimuli sociaux et émotionnels. *Psychologie Française*, *50*(1), 167-180.

Droit-Volet, S. (2005). Le temps psychologique. *Psychologie Française*, *50*(1), 5-6.

Droit-Volet, S. & Gil, S. (2009). The time-emotion paradox. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1525), 1943-1953.

Eagleman, D. M. (2008). Human time perception and its illusions. *Current Opinion in Neurobiology*, *18*(2), 131-136.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, *84*(3), 279-325.

Gil, S. & Droit-Volet, S. (2011a). How do emotional facial expressions influence our perception of time? *Emotion*, 1-12.

Gil, S. & Droit-Volet, S. (2011b). "Time flies in the presence of angry faces"… depending on the temporal task used! *Acta Psychologica*, *136*(3), 354-362.

Gil, S., Rousset, S., & Droit-Volet, S. (2009). How liked and disliked foods affect time perception. *Emotion (Washington, D.C.)*, *9*(4), 457-463.

Grondin, S. (2001). From physical time to the first and second moments of psychological time. *Psychological Bulletin 127*(1), 22-44.

Rousset, S., Deiss, V., Juillard, E., Schlich, P., & Droit-Volet, S. (2005). Emotions generated by meat and other food products in women. *British Journal of Nutrition, 94*, 1–12.

SU.VI.MAX (1994): Portions alimentaires. Manuel photo pour l'estimation des quantités. Paris, France: Polytechnica Editions.

Acknowledgement. This research was made possible by a research grant from the Social Sciences and Humanities Research Council of Canada awarded to SG.

A Category-Order Effect?:
Sub-Categorical Properties of Stimuli Determine a Categorical Effect

Jordan Richard Schoenherr[1], Robert Thomson[2], Grant McGee[1], Guy Lacroix[1]
[1]Department of Psychology, Carleton University
1125 Colonel By Drive, Ottawa, ON K1S5B6 Canada
[2]Department of Psychology, Carnegie Melon University
5000 Forbes Avenue, Pittsburgh, PA 15213 United States of America
Psychophysics.Lab@gmail.com, ThomsonR@andrew.cmu.edu, Guy_Lacroix@carleton.ca

Abstract

The category-order effect (COE) is observed when the categorical properties of items within the first half of a given list affect recall performance in an immediate serial-recall task. The present study examines whether this recall advantage is a consequence of categorical properties (e.g., semantic-relatedness and category set-size) or whether the advantage is due to sub-categorical properties (e.g., orthographic similarity and word frequency). Participants were presented with numeric stimuli and nouns from a variety of semantic categories while their orthography (Experiment 1) and frequency (Experiment 2) were systematically manipulated. The results suggest that a large portion of the COE can be attributed to a reduction of the detection threshold associated with the sub-categorical properties of the items.

In serial recall tasks, item information activates both long-term (i.e., item information) and short-term (i.e., order information; Healy, 1974; Lashley, 1951; Nairne & Kelley, 2004; Schoenherr & Thomson, 2008) memory components. These components affect the extent to which participants can rehearse and retrieve to-be-recalled items. In the present study, we examined the contribution of the long-term memory by varying stimulus properties associated with category membership (e.g., semantic category) as well as those that are associated with individual stimuli (e.g., word frequency).

Brooks and Watkins (1990) referred to the dependency of recall performance on the order in which items from a particular category are presented in a list as the *category-order effect* (or COE). Both Brooks and Watkins (1990) and Greene and Lasek (1994) observed that list recall was enhanced when numbers are presented in a list prior to words (Young & Supa, 1941) and when related words precede unrelated words. Greene and Lasek thus concluded that the COE effect arises when a small, homogenous category is presented before a large, heterogeneous category because items in the initial positions of a list are likely to be more strongly activated in long-term memory.

The full-list immediate recall literature suggests that semantic similarity could help explain the COE. For instance, Crowder (1979) presented participants with 10 item lists consisting of similar or dissimilar words. He observed enhance recall for similar words relative to dissimilar ones (see also Saint-Aubin, Ouellette, & Poirier, 2005). Moreover, research has also found that the probability of correctly recalling an item is inversely proportional to the number of associates for that item in a given list (for a review, see Nelson, 1989). A straightforward explanation of these finding is that once a word is presented, it activates both its representation and similar stimuli (with overlapping semantic or contextual information) in long-term memory. This phenomenon is known as the fan effect (Anderson, 1974; Anderson & Reder, 1999). As activation increases, semantically related units also become active thereby becoming candidates for recall, creating proportionally higher interference for items with a greater category size or items seen in more contexts than items with a smaller category size (West, Pyke, Rutledge-Taylor, & Lang, 2011). Some earlier evidence of the fan effect is presented in the work of Crannell and Parrish (1957). They examined the effect of set sizes and semantic categories on immediate serial recall. More specifically, participants were asked to remember sets consisting of digits (1-9; set size = 9), letters (from the full set of letters), letters (from a limited set of letters 'a' to 'i'), three-letter words (from a set comprised of 286 members), and three-letter words (from a set comprised of 9 members). Overall, Crannell and Parrish found that digits led to the highest recall and that letters were recalled more accurately than words. Thus, it appears that the category set size explains only part of the advantage for digits relative to letters and words in general and that another stimulus property (categorical or sub-categorical) stored in long-term memory that contributes to the recall advantage.

Despite the evidence for recall facilitation resulting from the availability information stored within long-term memory, it is not necessarily the case that this information need be categorical (West et al., 2011). Sub-categorical information can be used as the basis for grouping exemplars. One of the most robust findings in the literature on immediate serial recall is that the frequency of words in a language corpus is negatively correlated with their response threshold (Howes & Solomon, 1951). For instance, McGinnies, Comer, and Lacey (1952) showed that word length affects performance independently of word frequency (see also Postman & Adis-Castro, 1957). This finding is of considerable importance for COE experiments considering that word stimuli (e.g., goose, dog, sheep, ox) have more characters than digit stimuli (e.g., 1, 6, 9) which has previously been shown to affect recall (Cowan et al., 1992). Moreover, studies have also found that orthographic properties affect recall performance. For example, it has been found that lowercase words are reported more accurately than uppercase words (e.g., Perri et al., 1996; Jordan, Redwood, & Patching, 2003; cf. Smith et al., 1969). Given that words are reported in lowercase type set and numbers are presented as Indian-Arabic numerals, this difference may enhance recall performance resulting in an additional release from proactive interference solely from orthographic properties.

Present Study

Previous studies of the category-order effect and related phenomena have frequently used number-word lists as stimuli (Brooks & Watkins, 1990; Greene & Lasek, 1994; Young & Supa, 1941). Although this paradigm has been viewed as contrasting a small, homogeneous set of items (i.e., numbers) against a larger, heterogeneous set of items (i.e., animals), several properties of the stimuli prohibit such a direct interpretation. First, number stimuli have been presented as digits resulting in a decreased load during visual encoding of a single item (e.g., 4) relative to an equivalent word (e.g., four). Second, number stimuli have a higher frequency than the word stimuli used in the recall lists. Examining these values reveals that within the Brysbaert and New corpus[1] there is a greater occurrence for the number words ($WF_{BN} = 4.10$) than animal words ($WF_{BN} = 3.12$) used in previous experiment (e.g., Young & Supa, 1941).Given the potential methodological confounds of previous studies, the present experiment uses an immediate serial recall task to assess the effect of item and order information on recall performance. Lists of high-

---

[1] Word counts retrieved Novermber, 20, 2011. WF_{BN} is given by adding one and taking the log10 of the SUBTLx word count.

and low-frequency words were created from four categories based on word norms and these were paired with high- (e.g., one and two) and low-frequency (e.g., twenty and thirty) numbers printed as words. These manipulations allow us to directly test the effect of a sub-categorical property (e.g., word frequency) and a categorical property (e.g., category nouns and numbers).

In Experiment 1, we sought to replicate the category-order effect with the materials used by Brooks and Watkins (1990) while also examining the effects of orthographic properties on encoding. Experiment 2 examined the role of word frequency. If the COE is a product of the activation of information in LTM, then orthographic dissimilarity and differences in word frequency between list halves should increase recall performance due to sub-categorical rather than categorical properties of the stimuli.

Experiment 1

Method

Twenty-three Carleton University students participated in the study for course credit. Format congruency between noun and number stimuli was manipulated. Stimuli consisted of monosyllabic numbers and animal names (nouns) used in previous studies of the COE (Brooks & Watkins, 1990; Greene & Lasek, 1994; Young & Supa, 1941). In the incongruent condition, we partially replicated the conditions used in previous studies by using word lists with number stimuli presented as digits and nouns (e.g., dog, ox) presented in lowercase letters as well as presenting digits with uppercase letters. In the congruent condition, both nouns and numbers were presented as words written in either uppercase or lowercase letters. As in previous studies, participants were either required to recall items in forward or backward order.

Participants were told that an eight-item sequence of four letters and four numbers would be presented on the computer monitor. Each participant was provided with an answer sheet and instructed to write down the memory stimuli after a response cue indicating the direction of recall, either “FORWARD" or “BACKWARD”. The response cue followed a 250 ms inter-stimulus interval that occurred after the stimuli were presented. If the cue indicated forward, participants would write down the stimulus in the order it was presented in. Alternatively, if the cue indicated backward, participants would be required to respond with the order of the categories reversed while preserving the order of the items within the category. Instructions emphasized both speed and accuracy. There were 20 training trials and 48 experimental trials.

Results and Discussion

A repeated-measures ANOVA was conducted on proportion correct examining the within-subjects variable of category-order (number-word vs. word-number) and recall order (forward vs. backward) as a between-subjects measures. In an additional analysis, we restructured the data to examine the effect of format congruency (i.e., list halves with same or different formats).

Replicating the findings of earlier studies, we obtained a COE, $F(1, 21) = 5.704$, $MSE = .013$, $p = .026$. Participants recalled more items in Number-Word lists ($M = .6812$, SD = .1806) than in Word-Number lists ($M = .6403$, $SD = .1706$). This result suggests that both item and order information contribute to the successful recall of stimuli. Importantly, we also observed a significant effect of number format, $F(1, 21) = 18.783$, $MSE = .006$, $p < .001$. Supporting our hypothesis that encoding fluency affected recall, we found that participants recalled more items in number lists written as digits ($M = .6863$) lists than number lists written as words ($M = .6352$).

This suggests that the COE might be influenced by sub-categorical properties of the stimuli such as orthography rather than solely categorical information.

We additionally wanted to determine whether orthographically similar (e.g., all uppercase or lowercase words v. nouns and number digits) items of the stimuli affected recall. Restructuring the data to examine the effect of stimuli congruency within a list, we found that list half dissimilarity affected recall, *F*(1, 21) = 18.783, *MSE* = .003, *p* < .001. We observed that when two lists halves were presented in different format, (*M* = .686) greater recall was observed then for those in the same format (*M* = .634).

Experiment 2

Method

Twenty-two Carleton University students participated in the study for course credit. Experiment 2 replicated the procedure of Experiment 1. For Experiment 2, stimuli sets were created by selected to create two sets of high- and low-frequency words to exclude the possibility that one semantic category was more salient than another. One high frequency set consisted of terms pertaining to relatives (e.g., son, mom, father) whereas the other consisted of terms pertaining to units of time (e.g., day, month, second). One low frequency set consisted of colour terms (e.g., tan, aqua, orange) whereas the other consisted of carpenter's tools (e.g., nail, wrench, pliers). This resulted in high- (*M* = 4.10) and low-frequency (*M*= 2.62) words and High- (*M* = 4.10) and low-frequency (*M*= 2.64) numbers were used as stimuli. All sets were matched for mean word length. Participants were randomly assigned to the category set conditions.

Results and Discussion

Data from one participants were removed prior to analysis for failing to conform to task demands. A repeated-measures ANOVA was conducted on the proportion of items recalled with category-order (word-number vs. number-word), recall order (forward vs. backward), word frequency (high vs. low), and number frequency (high vs. low) without analyzing stimulus set.

Table 1. Mean proportion recalled and standard error in forward and backward orders for high and low word frequency conditions.

Recall Order	Word Frequency	Proportion Recall
Forward	High	.695 (.053)
	Low	.662 (.052)
Backward	High	.623 (.062)
	Low	.630 (.061)

An examination of the effect of category-order alone did not reveal any significant results, *F*(1,17) = 1.007, *MSE* = .009, *p* = .33. This could suggest that the category-order effect might in fact be a result of sub-categorical properties that were not controlled for in previous studies. Supporting this hypothesis, we obtained a significant interaction between word frequency and recall order, *F*(1,17) = 5.135, *MSE* = .003, *p* = .037. Table 1 demonstrates that recall performance was highest for forward recall with this effect increasing for lists of high frequency words. This suggests that items that have numerous traces stored in long-term memory benefit from early activation as a result of presentation order.

Further evidence for the effect of sub-categorical properties was also observed in the main effect of number frequency, $F(1,17) = 5.135$, $MSE = .003$, $p = .037$. More items were recalled in lists containing high-frequency numbers ($M = .707$) relative to low-frequency numbers ($M = .598$) again suggesting that the number of memory traces is a primary determinant of recall performance in this task.

General Discussion

The results of two experiments revealed the conditions in which item and order information interact to increase recall performance in serial recall tasks in which the items belong to two different categories. In Experiment 1, we replicated the findings of previous experiments. More items were recalled from lists which presented number items prior to words (Brooks & Watkins, 1990; Greene & Lasek, 1994; Young & Supa, 1941). The original interpretation of the COE attributed this finding to the categorical properties of the number and word stimuli: number stimuli were drawn from a smaller, more homogenous category than word stimuli. Experiment 1, however, extended these results. It was alsoobserved that number format (digit and word) and orthographic properties of list halves also contribute to the COE. In general, the greater the orthographic differences there were between the lists halves, the greater the increase in recall performance.

Having identified the importance of orthographic properties to encoding during the task, Experiment 2 examined the effect of another sub-categorical property: stimulus frequency. When both word and number frequency were controlled, we failed to observe a COE. This result is not surprising given that the earliest evidence provided for the COE by Brooks and Watkins (1990) were the results of Watkins' (1977) study of list halves that varied in word frequency. Although it might not be the case that all COEs are the result of stimulus frequency, those associated with numeric stimuli used in the present study and similar studies do appear to be a result of frequency effects reducing the detection thresholds for the stimulus. Whereas it could be argued that the repeated exposure of stimuli could be construed as a categorical property due to continuous association of items within memory, it seems more reasonable to classify these memory traces properties as sub-categorical properties of the stimuli.

Conclusions

The present study replicated the COE of Brooks and Watkins (1990) and Greene and Lasek (1994). It is assumed to arise from an interaction between categorical properties of stimuli and the order in which they were presented. When participants are presented with stimuli, a trace is created in short-term memory. Information associated with those stimuli is activated in long-term memory. When items share category membership, the associations between items in a list generally enhance recall performance (Saint-Aubin et al., 2005). However, when categories are large, the resultant spread of activation to other category members creates interference (Nelson, 1989). Thus, the COE was thought to arise when items from a smaller, relatively homogenous category preceded a larger, relatively heterogeneous category. The results of our experiments, however, lead us to question this characterization of the COE. Namely, although order and item information do contribute to recall performance, and that categorical properties of the stimuli likely affect recall performance, the initial detection threshold of the stimuli appears to account for more recall performance once it has been controlled. This finding also has implications for studies in that perceptual effects appear to contribute more to recall performance than knowledge effects.

References

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, 5, 451-474.

Anderson, J. R. & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, 128, 186-197.

Brooks, J. O., & Watkins, M. J. (1990). Further evidence of the intricacy of memory span. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 1143-1141.

Cowan, N., Day, L., Saults, J. S., Keller, T. A., Johnson, T., & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language, 31*, 1-17.

Crannel, C. W., & Parrish, J. M. (1957). A comparison of immediate memory span for digits, letters, and words. *Journal of Psychology*, *44*, 319-327.

Crowder, R.G. (1979). Similarity and order in memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 13, pp. 319-353). New York: Academic Press.

Greene, R. L., & Lasek, R. (1994). Category-order effects in memory span. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1391-1395.

Healy, A. F. (1974). Separating item from order information in short-term memory. *Journal of Verbal Learning & Verbal Behavior*, *13*, 644-655.

Howes, D. H., & Solomon, R. L. (1951). Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology, 41*, 401-410.

Jordan, T. R., Redwood, M., & Patching, G. R. (2003). Effects of form familiarity on perception of words, pseudowords and nonwords in the two cerebral hemispheres. *Journal of Cognitive Neuroscience, 15*, 537-548.

Lashley, K.S. (1951). The problem of serial order in behaviour. In L. A. Jefress (Ed.), *Cerebral Mechanisms in Behaviour: the Hixon Symposium*. Wiley.

McGinnies, E., Comer, P. B., & Lacey, O. L. (1952). Visual-recognition thresholds as a function of word length and word frequency. *Journal of Experimental Psychology*, *44*, 65-69.

Nairne, J. S., & Kelley, M. R. (2004). Separating item and order information through process dissociation. *Journal of Memory and Language*, *50*, 113-133.

Nelson, D. L., Keelean, P. D., & Negrao, M. (1989). Word-fragment cuing: The lexical search hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 388-397.

Perri, R., Bartolomeo, P., & Silveri, M. C. (1996). Letter dyslexia in a letter-by-letter reader. *Brain and Language*, *53*, 390-407.

Postman, L., & Adis-Castro, G. (1957). Psychophysical methods in the study of word recognition. *Science, 125*, 193-194.

Schoenherr, J. R. & Thomson, R. (2008). Category Properties and the Category-Order Effect. *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*, Washington, USA.

Smith, F., Lott, D., & Cronnell, B. (1969). The effect of type size and case alternation on word identification. *The American Journal of Psychology*, *82*, 248-253.

Saint-Aubin, J., Ouellette, D., & Poirier, M. (2005). Semantic similarity and immediate serial recall: Is there an effect on all trials? *Psychonomic Bulletin & Review*, *12*(1), 171-177.

Watkins, M. J. (1977). The intricacy of memory span. *Memory & Cognition, 5*, 529-534.

West, R. L., Pyke, A. A., Rutledge-Taylor, M. F., & Lang, H. (2010). Interference and ACT-R: New evidence from the fan effect. In D. D. Salvucci & G. Gunzelmann (Eds.), *Proceedings of the 10th International Conference on Cognitive Modeling* (pp. 277-281). Philadelphia, PA: Drexel University.

Young, C. W., & Supa, M. (1941). Mnemic inhibition as a factor in the limitation of the memory span. *The American Journal of Psychology*, *54*, 546-552.

Dissociating Categorization Learning System:
Overconfidence and Variations Learning Rates in Accuracy and Confidence Reports

Jordan Richard Schoenherr and Guy Lacroix
psychophysics.lab@gmail.com, guy_lacroix@carleton.ca
Department of Psychology, Carleton University
1125 Colonel By Drive, Ottawa, ON K1S5B6 Canada

Abstract

In examination of dual-process models of categorizations, research has typically focussed on the manner in which categorization responses change over time. However, one of the basic assumptions of a prominent dual-process account (COVIS) is that an explicit learning system dominates initial stages of training whereas an implicit learning system dominates later stages of training. In three experiments, we consider the utility of using subjective measures of performance (i.e., confidence reports) to continuously sample from a participant's explicit representation of the category structure while also examining changes in these reports over the course of training. The results of an examination of learning rates and the block at which participants reached a performance asymptote support multiple processes and representations accounts of categorization.

Dual-process models of categorization assume that information is processed by and represented in independent cognitive systems. These models have received support from a variety of sources including experimental studies, connectionist simulations, computational models, and neuroimaging studies (For a review, see Ashby & O'Brien, 2005). Several comprehensive dual-process models have been proposed (e.g., Erickson & Kruschke, 1998; Nosofsky, Gluck, Palmeri, McKinley, & Glauthier, 1994). The present study, however, will focus on a model based on Ashby, Alfonso-Reese, Turken, and Waldron, (1998) competition between verbal and implicit systems (COVIS). This model assumes that learning within the implicit system is dependent on feedback, and can be modelled using a multidimensional variant of single-detection theory (SDT) referred to as general recognition theory (GRT; Ashby & Townsend, 1986) and an explicit system that generates, tests, and modifies hypotheses of low-dimensional category structures.

COVIS presents two assumptions that are central to the present study. First, categorization decisions are made with a criterion, or category boundary (e.g., Ashby & Gott, 1988). In the context of the GRT, stimuli are assumed to have random perceptual effects defined by a joint probability distribution for each of their stimulus dimensions. With the provision of feedback, the category boundary divides separable or integral stimulus dimensions into discrete regions of a categorical space. If a stimulus consists of values along a dimension greater than those specified by the criterion, it is assigned to one category. If the values are less than that specified by the criterion, it is assigned to another category. Using curve fitting, Ashby and colleagues have demonstrated that by the end of training, participants performance is well-described by an optimal classifier model that employs a category boundary.

The second critical feature of COVIS is that the dominance of a given category learning system is dependent on the stage of learning and the structure of the category that has been acquired (Ashby & O'Brien, 2005). During initial stages of learning, categorization is assumed to be dominated by an explicit hypothesis-testing system. This system categorizes stimuli by identifying and testing one-dimension (1D) rule-based representations using executive resources and working memory. In parallel with the explicit learning process, a feedback-driven procedural-learning process associates a response category with a given region of perceptual space through stimulus-response mapping permitting the retention of a number stimulus dimensions. As the number of implicit memory traces increases over the course of learning, the activation level of the categorization response within the implicit system eventually exceeds that of the explicit system. Thereafter, the implicit system dominates response selection. When no feedback is provided, or when it is delayed by 2500 ms from the offset of a stimulus, participants have difficulty acquiring information-integration categories (e.g., Maddox, Ashby, & Bohil, 2003). Moreover, as the implicit system dominates response selection, response times decrease due to automaticity with longer response times that are found early in training suggesting competition between the implicit and explicit systems.

Subjective measures of performance such as confidence reports were amongst the earliest tools used in the context of experimental psychology and psychophysics to assess difference between awareness and performance (for a review, see Baranski & Petrusic, 1998). Retrospective confidence reports are typically obtained by having an individual assign a subjective probability (e.g., 60%) to the belief that they have provided a correct response. The degree of correspondence between a participant's mean accuracy when assigning a subjective probability to a response is referred to as *subjective calibration* (e.g., Baranski & Petrusic, 1994). Perfect calibration requires that the proportion correct (e.g., 0.6) and mean confidence are equivalent (60%) whereas miscalibration such as over-/underconfidence represents a bias. Studies of perceptual discrimination and general knowledge (e.g., Baranski & Petrusic, 1994) as well as memory (e.g., Koriat, 1993) have observed systematic deviations in the correspondence between task accuracy and subjective probabilities. These deviations can be attributed to differences in the operations supporting primary decision response selection and confidence processing.

An important assumption of COVIS is that the primary decision is based on a multidimensional model of SDT (GRT) and that subjective confidence is determined from a direct-scaling of this evidence (Ashby et al., 1998). With this in mind, it is critical to note that although SDT-based models of confidence processing (e.g., Ferrel & McGooey, 1980), including that proposed by Ashby et al. (1998) are parsimonious, they cannot readily account for several robust findings in the confidence literature. First, the systematic deviations observed in confidence calibration suggest that a direct-scaling of primary decision evidence might not be the sole mechanism used to generate a confidence report (cf. Pleskac & Busemeyer, 2010). Supporting this possibility, studies have demonstrated that the calibration of subjective assessments of performance has been affected by sources of information other than that provided by the target stimulus (Busey, Tunnicliff, Loftus, & Loftus, 2000; Schoenherr, Leth-Steensen, & Petrusic, 2010) and requires additional operations associated with increased decision response time (DRT) when they are reported (e.g., Baranski & Petrusic, 1998; Schoenherr, 2009). Thus, the present study examines whether response accuracy and subjective confidence 1) differ between training blocks and learning rule, as well as whether 2) their learning rate changes depending on task requirements, and 3) the requirement of confidence increases DRTs.

For the purposes of this study we assume that the degree of correspondence between measures of accuracy and confidence can be used to infer the accessibility of representations and the underlying architecture of categorization processes during different stages of learning. First, when a performance asymptote is used, confidence should reach criterion prior to accuracy given the flexibility of the hypothesis-testing system and should exhibited a more rapid learning rate. Following from this, we additionally assume that participants should exhibit overconfidence when the category structure is readily verbalizable. Third, the requirement of confidence should also increase DRT if it constitutes a secondary process. Moreover, if the hypothesis-testing system and confidence share the same basis, automaticity of responses should occur more rapidly in the rule-based condition relative to the information-integration condition.

Experiment 1a

Method

Stimuli consisted of Gabor patches varying in terms of spatial frequency and orientation. Replicating the method of earlier studies (e.g., Zeithamova & Maddox, 2007), 40 Gabor patches were created for each category for the training phase using the randomization technique by randomly sampling values from two normal distributions. Stimulus values were rescaled into stimulus dimensions with spatial frequency given by $f = .25 + (x_1/50)$ and orientation given by $o = x2(\pi/500)$. Using these values, stimuli were generated with the Psychophysics Toolbox (Brainard, 1997) using MATLAB R2008 (MathWorks, Matick, MA) with an 85% performance asymptote. After a categorization response was provided and a confidence report was obtained, a feedback signal was presented to indicate a participant's accuracy in completing the task. Stimuli were presented to participants using E-Prime experimental software on a Dell Dimension desktop PC.

Procedure

The classification task procedure involved a training phase, consisting of 10 blocks of 40 replications per category, and a transfer phase consisting of 2 blocks of 40 replications per category. Participants were assigned to either the rule-based (RB) or information-integration (II) category structure. In Experiment 1a, participants were provided with both trial-to-trial and block feedback during the training phase. After feedback was provided, participants reported confidence on a 50 (guess) to 100 (certain) scale. In Experiment 1b, only trial-to-trial feedback was provided and in Experiment 2 trial-to-trial feedback was delayed by 2500 ms and the duration of the confidence report.

Figure 1a.

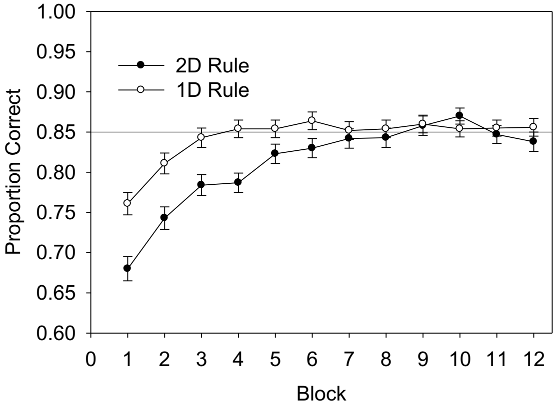


Figure 1b.

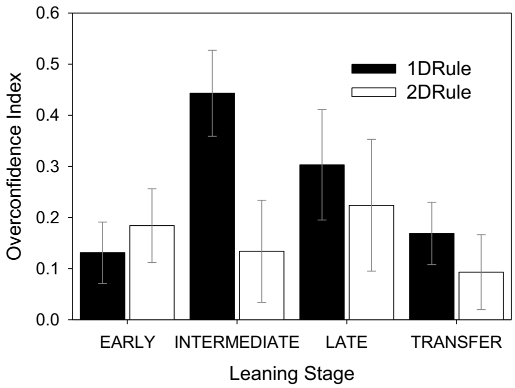


Figure 1. Categorization accuracy (Figure 1a) and overconfidence (Figure 1b) across learning and transfer blocks for 1D (white) and 2D rules (black).

Results

Proportion Correct. As demonstrated in Figure 1a, the results of categorization accuracy replicated earlier findings: 1D rules were learned in fewer blocks then 2D rules, $F(1, 83) = 6.317$, $MSE = .039$, $p = .014$, $\eta^2_p = .071$, and accuracy increased with the number of experimental blocks, $F(11, 913) = 49.167$, $MSE = .005$, $p < .001$, $\eta^2_p = .372$, as well as their interaction, , $F(11, 913) = 6.891$, $MSE = .005$, $p < .001$, $\eta^2_p = .077$. We also found that the requirement of confidence affected category learning as it interacted with block, $F(11, 913) = 2.093$, $MSE = .005$, $p = .052$, $\eta^2_p = .025$. Although the requirement of confidence initially produced reduced performance in the first block ($M = .703$, $SD = .140$) relative to no confidence ($M = .738$, $SD = .131$), participants who reported confidence in the transfer phase were more accurate ($M = .866$, $SD = .112$) then those who did not ($M = .829$, $SD = .103$).

Decision Response Time. Prior to conducted the ANOVA on decision response time (DRT), outliers three standard deviations above the mean were first removed. This accounted for 2.1 % of the total data. As in the accuracy results, participants response time decreased from early to later blocks of training, $F(11, 913) = 11.792$, $MSE = 335411$, $p < .001$, $\eta^2_p = .096$. Of considerable interest to the present study, we found the interaction of experimental block and confidence condition to be significant, $F(11, 913) = 4.600$, $MSE = 335411$, $p = .001$, $\eta^2_p = .039$. As is clear from Figure 3, longer response latencies were observed in conditions when confidence was required. Whereas DRT is relatively constant in the no confidence condition, two rapid rates of change are observed in the DRT when confidence is required between Block 1 and 2 as well as between Block 6 and 7 for the rule-based and information-integration conditions, respectively. This pattern of automaticity suggests that confidence reports required greater additional processing during the categorization response for information-integration category structure than for rule-based category structures.

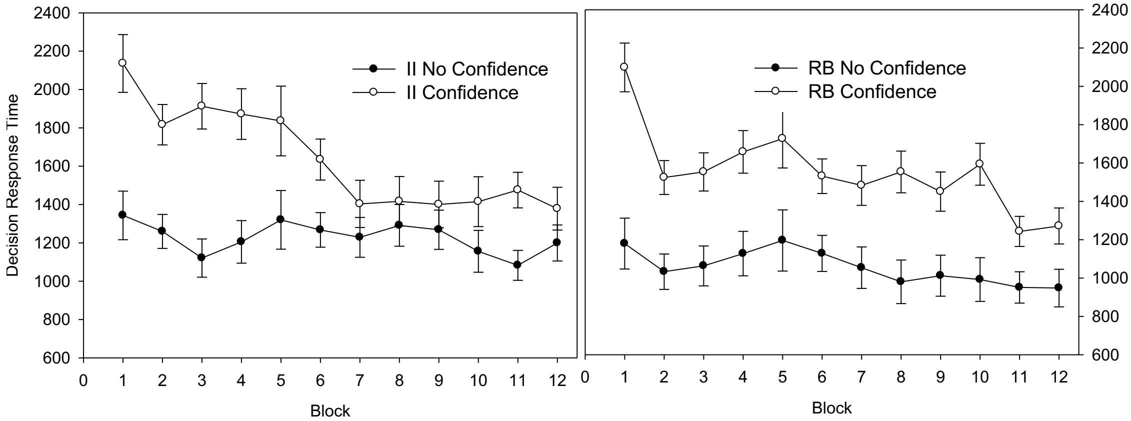


Figure 2. Decision response times with and without the requirement of confidence reports in the rule-based and information-integration learning conditions.

Confidence Reports. Overall, we found that the overconfidence bias differed across the learning phases (see Figure 1b), $F(1,77) = 8.842$, $MSE = .085$, $p = .004$, $\eta^2_p = .103$. As expected, learning phase was also found to interact with category structure, $F(1,77) = 4.539$, $MSE = .085$, $p = .036$, $\eta^2_p = .056$. As can be seen from Figure 4, overconfidence remained relatively constant in the information-integration condition suggesting that, in general, participants did not have access to the representation that guided their performance. In contrast, an increase in overconfidence was observed in intermediate phases of training in the rule-based condition. This finding suggests that once participants identified the one-dimensional rule, they expected to have continual improvements in performance.

Discussion

The results of our experiment replicate several earlier studies within categorization and confidence processing literature. Categorization performance was found to be affected by the nature of the category structure participants are required to learn: Participants who were required to learn the 1D category structure reached a performance asymptote earlier than those who were required to learn the 2D category structure. Moreover, we also observed that response latencies decreased in fewer blocks for those learning the 1D structure relative to the 2D category structure indicting that participants could more readily acquire a stimulus-response mapping for rule-based categories relative to information-integration categories. Findings such as these conform to the predictions of dual-process accounts of categorization such as COVIS (Ashby et al., 1998) and thereby allow us to interpret the results obtained from confidence reports.

Our analysis of confidence reports also provides evidence for a dual-process account. In the experiment conducted here, we observed increased overconfidence in intermediate phases of training for those participants learning a 1D category structure relative to those who learned the 2D category structure. In general, miscalibration observed here indicates that the representation used to report subjective confidence and that used to respond to exemplars were informed by different sources of information. Greater overconfidence suggests that the category structure that participants were explicitly aware of did not contain the stimulus variability evidenced in the distribution of exemplars. Although it is possible that during the process of rescaling primary decision accumulated evidence could have decayed, it is less clear how this could have resulted in the increases in confidence that lead to the overconfidence observed in our data. More plausibly, it would seem to be the case that different stimulus representations of the category structure were available to two categorization systems.

Table 1. Mean difference between confidence and accuracy for slope and performance asymptote functions across experimental conditions.

Condition	Experiment/Rule	Slope Difference	85% Asymptote Block
Block Feedback	EX1a (1D)	.007	0.23
	EX1a (2D)	-.012	0.45
No Block Feedback	EX1b (1D)	.019	0.12
	EX1b (2D)	.007	0.25
Delayed Trial Feedback	EX2 (1D)	.003	0.36
	EX2 (2D)	.006	0.42

Due to space limitations additional preliminary results of additional experiments could not be reported in great detail. Table 1 contains differences in growth rates for mean accuracy and mean confidence data for the two category structures with the removal of block feedback (Experiment 1b) and when feedback was delayed (Experiment 2). Logistic functions were fit for mean participants performance allowing us to obtain the rate of growth (slope) and a measure of the estimate at which point participants typically reached the performance asymptote (85% correct). As is clear from Table 1, there were generally greater observed positive differences between the slope and performance asymptote for accuracy and confidence functions in the 1D condition relative to the 2D when participants were provided with full feedback (Experiment 1a) or trial-to-trial feedback (Experiment 1b). This suggests that the category structure participants were explicitly aware of was acquired more rapidly than the representation used to inform categorization responses. Supporting this interpretation, differences in the time taken to reach the performance asymptote was much greater in the 2D condition relative to the 1D condition suggesting that participants estimated that they had reached 85% much earlier then they in fact did. Moreover, this difference was again greater when feedback was delayed. Taken together with the analysis of accuracy and confidence indices, we take these findings to suggest two ostensibly distinct category learning and representation systems.

References

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, 105, 442-481.

Ashby, F. G., & O'Brien, J. B. (2005). Category learning and multiple memory systems. *Trends in Cognitive Sciences*, 9, 83-89.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *JEP: LMC*, 14, 33-53.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, 93, 154-179.

Baranski, J. V., & Petrusic, W. M. (1994). The Calibration and resolution of confidence in perceptual judgements. *Perception & Psychophysics*, 55, 412-428.

Baranski, J. V., & Petrusic, W. M. (1998). Probing the locus of confidence judgments: Experiments on the time to determine confidence. *JEP: HPP*, 24, 929-945.

Brainard, D. H. (1997). The Psychophysics Toolbox, *Spatial Vision* 10, 433-436.

Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin & Review*, 7, 26-48.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, 127, 107-140.

Ferrel, W. R., & McGoey, P. J. (1980). A model of calibration for subjective probabilities. *Organizational Behaviour and Human Performance*, *26*, 32-53.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, 100, 609-639.

Maddox, W. T., Ashby, F. G., & Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. *JEP: LMC*, 29, 650-662.

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition*, 22, 352-369.

Pleskac, T.J. and Busemeyer, J.R. (2010) Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychological Review*, 117, 864–901.

Schoenherr, J. R. (2009). *Mapping Internal Representations of Confidence onto Scales Varying in Range, Interval, and Number of Response Categories*. Unpublished Manuscript.

Schoenherr, J. R., Leth-Steensen, C., & Petrusic, W. M. (2010). Selective attention and subjective confidence calibration. *Attenion, Perception & Psychophysics*, 72, 353-368.

Zeithamova, D., & Maddox, W. T. (2007). The role of visuo-spatial and verbal working memory in perceptual category learning. *Memory & Cognition*, 35, 1380–1398.

TEMPORAL PERCEPTION OF ACOUSTICALLY PRESENTED TIME SERIES

Jiří Wackermann, Jakob Pacer, and Marc Wittmann
Dept. of Empirical and Analytical Psychophysics
Institute for Frontier Areas of Psychology and Mental Health
Freiburg i. Br., Germany
*<jw@igpp.de>

Abstract

Data from three experiments on serial perception of temporal intervals are reported. Sequences of short acoustic signals ('pips') separated by periods of silence were presented to the observers (N = 24, 16, 12). The inter-pip intervals were generated to form geometric or alternating series, while the modulus $1{+}\delta$ *of the stimulus series and the base duration* T_0 *(in the range from 1.1 to 6 seconds) were varied as independent parameters. The observers had to judge whether the pips sequences were accelerating, decelerating, or uniform, or to distinguish regular from irregular sequences. 'Intervals of subjective uniformity' (*ISU*) were obtained by fitting Gaussian psychometric functions to individual subjects' responses. Progression towards longer base durations (*$T_0 = 4.4$ *or 6 seconds) shifts the* ISU*s towards negative* δ*s, i. e., accelerating series. This finding is compatible with the phenomenon of 'subjective shortening' of past temporal intervals, accounted for by the lossy integration model of internal time representation.*

In a standard duration discrimination task, pairs of intervals marked by e. g. acoustic or visual stimuli are presented to the observer who indicates which member of the pair was perceived as longer or shorter (2AFC paradigm), or reports their subjective indifference (3AFC paradigm). In the serial variant of the task, sequences of intervals generated by a recursive rule are presented to the observer who has to categorize the series as a whole, according to a given instruction. The idea behind this procedural modification is that the 'subjective evidence' underlying the global judgment of the stimulus builds up sequentially from 'local' comparisons between subsequent intervals. Therefore, the sequential procedure should allow to study discrimination processes more efficiently than by collecting responses from pairwise comparisons. In the study reported here we investigated discrimination of acoustically presented intervals in the supra-second domain, resulting in judgments of 'acceleration' vs. 'deceleration', or 'regularity' vs. 'irregularity' as the global property of the stimulus.

Methods

Apparatus and stimuli

Stimuli were generated by a dedicated program running under BSD Unix on a portable iBook G4 (Apple Inc.) computer, with a pointing device ('mouse') and an external LCD monitor attached. The sound output of the computer was fed via an amplifier (Sony TA-FE310R) to a pair of headphones (Sennheiser HD 201) worn by the observers during the experimental session. The acoustic markers ('pips') of time intervals were sound signals with a frequency of 2000 Hz and a duration of 20 ms. The control program generated two types of stimuli: monotonically modulated time series (mode mono), or alternating time series (mode mach[1]), using the formulae

$$\begin{aligned} T_n &= (1+\delta)\,T_{n-1} && (\mathsf{mono}) \\ T_n &= (1+(-1)^n\,\delta/2)\,T_0 && (\mathsf{mach}) \end{aligned}$$

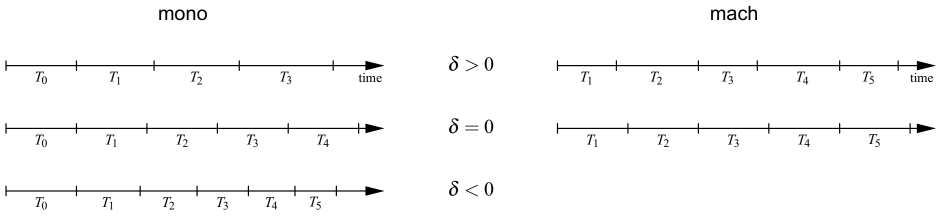


Figure 1. Monotonic (left) and alternating (right) pips sequences for different values of δ.

T_0 (base duration) and δ are two control parameters determining the temporal structure of the stimulus (Fig. 1), which were varied according to the experimental design (see below).

Procedures and designs

In each trial, the observer was listening to the pips series played through the headphones; thereafter a list of possible responses was displayed on the monitor, from which the observer had to choose. Duration of the pips series was limited to max. 1 minute, but the observer could interrupt the stimulus presentation by pressing a button, after which the response menu was displayed immediately. An exception from this was Exp. 3 where the number of presented intervals n_{P} was fixed by design. The 3AFC paradigm with response categories 'accelerating', 'uniform', and 'decelerating' was used for the monotonic series (mono). The 2AFC paradigm with response categories 'regular' and 'irregular' was used for the alternating series (mach).

Experiment 1: Twenty-four observers (12 women and 12 men, age range 21–33 years, mean age 25.8 years) participated in the experiment. Three blocks, each consisting of twelve trials, were run with different base durations in a fully permuted order, using $T_0 = 1.5, 3, 6$ s for a sub-group of twelve subjects, and $T_0 = 1.1, 2.2, 4.4$ s for the other twelve subjects. Each block consisted of two parts. In part 1, δ varied from -0.05 to $+0.05$ in steps of 0.02, following an interlaced up-and-down staircase scheme, with two repetitions for each δ value. A point of subjective uniformity (PSU) was roughly estimated from the data, and a new range of δs distributed symmetrically around the PSU with halved steps of 0.01 was used in part 2, following the same scheme. This two-phase procedure was designed to adapt the δ-sampling scheme to the subject's individual performance.

Experiment 2: Sixteen observers (8 women and 8 men, age range 21–29 years, mean age 24.4 years) participated in the experiment. Each session consisted of two blocks, one block with monotonic series (mode mono) and the other block with alternating series (mode mach). Two base durations were used in each block, $T_0 = 2.2$ or 4.4 s with nine subjects, and $T_0 = 3$ or 6 s with seven subjects, in a permuted order.[2] A mono block consisted of 26 trials, with δ varied from -0.06 to $+0.06$ in steps of 0.01, with two repetitions for each δ value. A mach block consisted of 22 trials, with δ varied from 0 to 0.25 in steps of 0.025 for base durations 2.2 s and 3 s, or a doubled range up to 0.5, and doubled steps of 0.05 for base durations 4.4 s and 6 s. These settings were based on a series of pilot experiments showing that the discrimination in the mach mode was definitely inferior to that in the mono mode. An up-and-down staircase δ-sampling scheme similar to Exp. 1 was used.

Experiment 3: Twelve observers (6 women and 6 men, age range 21–28 years, mean age 24.7 years) participated in the experiment. Unlike Exps. 1 and 2, the number of presented intervals was limited to $n_{\mathrm{P}} = 2$, 5, or 10. The observer had to listen to the entire sequence before

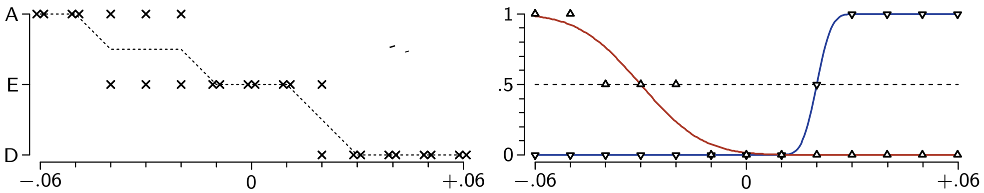


Figure 2. Transformation of a 3AFC dataset to two PMFs. Left: Individual responses (×) for δ varying from -0.06 to $+0.06$: A = accelerating, E = uniform, D = decelerating, dashed broken curve = central response tendency. Right: Relative frequencies of responses 'accelerating' (△) and 'decelerating' (▽), to which psychometric functions $\psi_{\mathrm{A}}(\delta)$ (red) and $\psi_{\mathrm{D}}(\delta)$ (blue) are fitted; the two curves cross the dashed line $P = 0.5$ at points $\delta = \theta_{\mathrm{A}}$ and $\delta = \theta_{\mathrm{D}}$, respectively. (Illustrative data from Experiment 2, mode mono, $T_0 = 3$ s.)

giving a response.[3] The three n_{P} values were combined factorially with two base durations, $T_0 = 3$ and 6 s, thus resulting in six blocks, each block consisting of 18 trials. For $n_{\mathrm{P}} = 5$ and 10, δ varied from -0.1 to $+0.1$ in steps of 0.025; for $n_{\mathrm{P}} = 2$, the δ range was adjusted to $[-0.4, +0.4]$, and the steps increased accordingly to 0.1. Again, these settings were based on exploratory pilot experiments. An up-and-down staircase δ-sampling scheme similar to Exp. 2 and Exp. 3 was used.

Data reduction

Separately for each subject and base duration T_0, data were sorted by δ, relative frequencies of response categories were calculated, and psychometric functions (PMF) were fitted to the frequencies using the maximum-likelihood method. For the 3AFC data (mode mono), probabilities of responses 'accelerated' or 'decelerated' as functions of δ (cf. Fig. 2) are given by

$$\psi_{\mathrm{A}}(\delta) = \Phi\left(\frac{\delta - \theta_{\mathrm{A}}}{\sigma_{\mathrm{A}}}\right), \qquad \psi_{\mathrm{D}}(\delta) = \Phi\left(\frac{\delta - \theta_{\mathrm{D}}}{\sigma_{\mathrm{D}}}\right).$$

For 2AFC data acquired in the mode mach, the probability of response 'regular' is

$$\psi_{\mathrm{R}}(\delta) = \Phi\left(\frac{\delta - \theta_{\mathrm{R}}}{\sigma_{\mathrm{R}}}\right)$$

Where Φ denotes the normal Gaussian CDF, parameters θ_{A} and θ_{D} delimit an 'interval of subjective uniformity' (ISU),[4] and θ_{R} determines a threshold for detection of irregularity.

Results

Of interest are primarily dependences of the ISUs on experimental conditions, i. e., base durations T_0, presentation modes (mono, mach) and number of presented intervals n_{P}.

Experiment 1

The ISU width $\theta_{\mathrm{D}} - \theta_{\mathrm{A}}$ varied from ~ 0.023 at shortest base durations, 1.1 and 1.5 s, to 0.052 for the longest, $T_0 = 6$ s, whereas the ISU midpoints $\tilde{\theta} = (\theta_{\mathrm{A}} + \theta_{\mathrm{D}})/2$ shifted from $+0.005$ to -0.004 (Fig. 3a). A closer look at the data reveals that these effects are due to a significant and systematic shift of the acceleration threshold towards negative values with increasing T_0 (Fig. 3b), verified by intraindividual pairwise t tests (all with $df = 11$): 1.1 vs. 2.2 s: $t = 1.94$

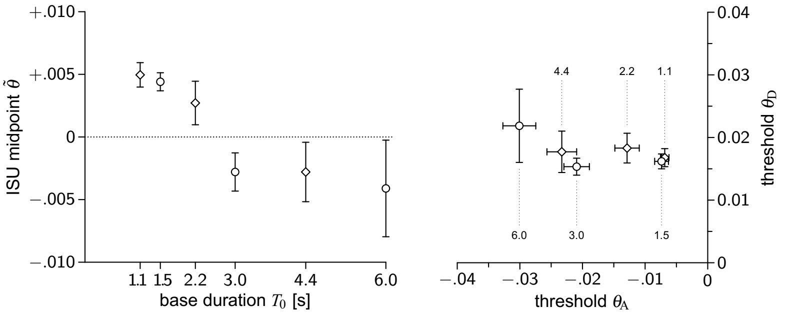


Figure 3. Dependence of ISU on base duration T_0 in Experiment 1. Left: Mean ISU midpoint $\bar{\tilde{\theta}}$ plotted as a function of base duration T_0. Right: Interrelation between mean thresholds $\bar{\theta}_A$ and $\bar{\theta}_D$ across different base duration T_0 (indicated by tiny numbers and dotted lines). Group means ±1 SEM are shown ($N = 12$ for each data point). Two subsets of different T_0 are distinguished by graphic symbols ◇ and ○.

($P \approx 0.082$), 1.5 vs. 3 s: $t = 5.00$ ($P < .005$), 2.2 vs. 4.4 s,: $t = 3.57$ ($P < .02$), 3 vs. 6 s: $t = 2.62$ ($P < .05$). On the other hand, the 'deceleration threshold' θ_D did not change significantly with T_0, only varied around $\sim$0.018. This 'dissociation' between acceleration and deceleration threshold is also supported by insignificant pairwise correlations between θ_A and θ_D.

Interestingly, three subjects failed to detect an acceleration within the entire range of δ (one subject at $T_0 = 4.4$ s, two subjects at $T_0 = 6.0$ s). These cases can be interpreted as extreme manifestations of the 'θ_A traveling' effect.[5]

Experiment 2

The results for the mono mode only partially confirm those from Exp. 1. Comparisons between $T_0 = 2.2$ and 4.4 s show a negative shift of *both* θ_A ($t = 2.15$, $df = 8$, $P \approx .067$) and θ_D ($t = 2.60$, $df = 8$, $P < .05$), resulting in a significant shift of the midpoint $\tilde{\theta}$ of the *entire* ISU ($t = 3.34$, $df = 8$, $P \approx .01$). Comparisons between $T_0 = 3$ and 6 s show also negative but non-significant shifts of the ISU.

As for the mach mode, mean regularity thresholds $\bar{\theta}_R$ were 0.221 for $T_0 = 2.2$ s, and slightly lower for longer base durations: 0.141 ($T_0 = 3$ s), 0.188 ($T_0 = 4.4$ s), and 0.163 ($T_0 = 6$ s). These values, which are by almost one order of magnitude higher than mean θ_A or θ_D, indicate that recognition of deviation from regularity is considerably more difficult for non-directed than directed (monotonic) interval series. Intraindividual pairwise comparisons between shorter and longer base durations (2.2 vs. 4.4 s, or 3 vs. 6 s) do not reveal any significant differences. No significant correlations were found between θ_A and θ_R, or θ_D and θ_R. There seems to be no strong relation between duration discrimination in the two modes, mono and mach.[6]

Experiment 3

The mean ISU width $\theta_D - \theta_A$ decreased from 0.388 for $n_P = 2$ via 0.118 for $n_P = 5$ to 0.084 for $n_P = 10$ (average across both base durations, 3 and 6 s).[7] This effect obviously reflects a reduction of the observer's uncertainty with increasing number of perceived intervals. For all three n_P conditions, the ISU midpoints $\tilde{\theta}$ shifted from positive values at $T_0 = 3$ s to negative values at $T_0 = 6$ s, reflecting simultaneous shifts of *both* θ_A and θ_D. In spite of this common pattern, ISU shifts are statistically significant only for $n_P = 5$: for θ_A, $t = 2.80$ ($P < .02$), for θ_D, $t = 3.16$ ($P < .01$), for $\tilde{\theta}$, $t = 3.48$ ($P < .01$) (all t tests have 11 df). The ISU width did not change significantly with base duration.

Discussion

In three experiments we studied serial perception of ‘empty’ temporal intervals marked by acoustic signals (‘pips’). There is a large volume of literature on perception of empty vs. filled intervals on the one hand, and on perception of rhythm on the other hand, which cannot be discussed here for space limitations. We only point out that the major focus of our study is on the transition from the circa-second to the supra-second domain, where serial perception changes qualitatively, from rhythmic patterns to series of discrete events.

We assume that detection of a global property of a time series results from accumulating a series of comparisons between subsequent intervals. Two types of stimuli, geometric series (mono), or alternating series (mach) were used, for which the following relations hold:

$$\frac{T_{n+1} - T_n}{T_n} = \delta \quad \textsf{(mono)}, \qquad \frac{T_{n+1} - T_n}{T_n} \approx \pm\delta \quad \textsf{(mach)},$$

This constancy of Weberian ratios between durations of subsequent intervals[8] suggests a more-or-less uniform accumulation of cognitive evidence resulting in the observer’s response. Elaboration of a more specific mathematical model of the process is a task for the future.

Our use of a three alternatives response scheme (3AFC) in trials with monotonic time series was based mainly on observations from early pilot experiments: in some subjects the region of subjective uncertainty between ‘accelerating’ and ‘decelerating’ may be quite wide, so it is favourable to provide a neutral alternative (‘uniform’, ‘neither–nor’). We prefered not to force the participants to give responses that do not match their subjective perceptual experience. The 3AFC paradigm turned out to be a fortunate choice, as it allowed us to analyze temporal discrimination in terms of two parameters, θ_A and θ_D, delimiting the interval of subjective uniformity (ISU).

The main result of this study, suggested by Exp. 1 and at least partly confirmed by results of Exps. 2 and 3, is the systematic shift of the ISU with increasing base duration T_0 in monotonic time series. Whether the dissociation between θ_A and θ_D, seen in Exp. 1, is a rule or an exception from the ISU shift as a whole (Exp. 2, 3) is still an open question. Provisionally, the term ‘ISU traveling effect’ used in this paper covers both cases. Importantly, results from Exp. 1 suggest that the ISU shift is not a linear or linear-like function of T_0, but rather a steep transition occurring between $T_0 = 2.2$ s and 3 s (cf. Fig. 3a). This observation is in line with evidence showing that intervals up to 2 to 3 s are processed differently than longer intervals. Temporal integration of events up to 2 to 3 s has been reported in many qualitatively different experiments in perception and action (Pöppel, 1997; Wackermann, 2007; Wittmann, 2011). Temporal intervals exceeding this window of the ‘psychological present’ require additional short-term memory processes since a temporal interval covering multiple seconds has to be re-constructed from memory. The ISU shift starting at intervals > 2 s as found in our data adds to the idea of different processing mechanisms for intervals within and beyond the ‘psychological present’.

The ISU ‘traveling effect’ observed for longer durations in the supra-second domain implies an increasing overlap between the ISU and the negative part of the δ continuum; in other words, a region of ‘illusory perception’ where objectively accelerated series are perceived as uniform. This phenomenon can be explained, at least qualitatively, by the lossy integration model of internal time representation (Sysoeva et al., 2011; Wackermann, 2011), also known as ‘dual klepsydra model’ (DKM) (Wackermann & Ehm, 2006). The model accounts for the progressive shortening of reproduction responses as well as for the asymmetry in duration discrimination between pairs of intervals (Wackermann & Späti, 2006). Applying the model to series of $n > 2$

intervals, it can be shown (Wackermann, 2008) that the series *must* accelerate to be perceived as a sequence of equal periods.[9] The effect reported here thus supports the evidence for the lossy integration model.

Notes

[1] Mach (1865) was the first to use alternating acoustic series in a study of temporal discrimination.

[2] Due to an operator's error, this distribution deviates from the intended balanced design, 2×8 subjects for each pair of base durations. However, the error does not affect the subsequent statistics.

[3] For $n_P = 2$, the observer's task is reduced to a comparison between two empty intervals marked by three pips. Response alternatives 'the second interval was shorter' or 'the second interval was longer' were used and coded in data as 'accelerating' or 'decelerating', respectively.

[4] Note that an ISU $\equiv [\theta_A, \theta_D]$ is truly an interval in the domain of real numbers; this in contrast to the usual parlance of time research, referring to durations, i. e. single values, as 'intervals'.

[5] For the statistics, the missing θ_A values were replaced by the group mean across all remaining subjects for the same base duration. This is a super-conservative estimate, in fact somewhat counter-acting the observed effect which, however, still remains significant.

[6] Note, however, that the numbers of observations in each of the analyzed data subsets are too small to allow reliable inferences.

[7] Two subjects yielded extremely flat PMFs with $T_0 = 6$ s and $n_P = 2$ or 5, respectively, so that the ISU could not be estimated reliably. Similarly as in Exp. 1, the missing θ_A and θ_D values were replaced by the group means across all remaining subjects for the same conditions.

[8] This, of course, does not imply that we assume duration discrimination to be strictly Weberian.

[9] The recursion rule for 'klepsydraically uniform' time series differs from the linear recursion used in our experiments, so our data do not allow a straightforward estimation of the DKM parameter. We reserve this problem for a later treatment.

[*] Thanks are due to Oksana Gutina for conducting part of the experiments and for general assistance.

References

Mach, E. (1865). Untersuchungen über den Zeitsinn des Ohres. *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften (Wien), math.-naturwiss. Klasse*, *51*, 135–150.

Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends in Cognitive Science*, *1*, 56–61.

Sysoeva, O. V., Wittmann, M., & Wackermann, J. (2011). Neural representation of temporal duration: Coherent findings obtained with the "lossy integration" model. *Frontiers in Integrative Neuroscience*, *5*, 37.

Wackermann, J. (2007). Inner and outer horizons of time experience. *Spanish Journal of Psychology*, *10*, 20–32.

Wackermann, J. (2008). Measure of time: A meeting point of psychophysics and fundamental physics. *Mind and Matter*, *6(1)*, 9–50.

Wackermann, J. (2011). On clocks, models and metaphors: Understanding the klepsydra model. In: Vatakis, A., Esposito, A., Giagkou, M., Cummins, F., & Papadelis, G. (Eds.) *Multidisciplinary Aspects of Time and Time Perception* (Berlin/Heidelberg: Springer) pp. 246–257.

Wackermann, J. & Ehm, W. (2006). The dual klepsydra model of internal time representation and time reproduction. *Journal of Theoretical Biology*, *239*, 482–493.

Wackermann, J. & Späti, J. (2006). Asymmetry of the discrimination function for temporal durations in human subjects. *Acta Neurobiologiae Experimentalis*, *66*, 245–254.

Wittmann, M. (2011). Moments in time. *Frontiers in Integrative Neuroscience*, *5*, 66.

Can Pitch Be Measured On A Ratio Scale?

Florian Kattner and Wolfgang Ellermeier
Technische Universität Darmstadt
kattner@psychologie.tu-darmstadt.de

Abstract

Stevens' direct scaling methods have been widely used in basic and applied psychophysical research. They rest on the assumption that observers are capable of processing ratios on a subjective intensity scale. Axioms fundamental to this assumption have only recently been formulated, and empirically tested for a limited number of sensory continua, most notably loudness. The present study attempts to extend this line of research to the perception of pitch. N=13 participants were asked to adjust pitch intervals defined by pure tones in a frequency range between 264 and 699 Hz to specific ratios (e.g., 1/2 or 2/3) of a standard interval. Whereas most participants' adjustments were in accordance with the axiom of monotonicity, there were a few significant violations of commutativity and multiplicativity, particularly if the standard interval exceeded an octave. In contrast to what has been found for loudness, however, multiplicativity still held for most observers. Systematic effects of stimulus range and musical training appear to distinguish pitch scaling from other quantifiable sensations.

In a typical application of direct psychophysical scaling, the observer is asked to assign numbers to the intensities of sensations (Stevens, 1956, 1975). That is, the participant is either responding with a numeral to the intensity of stimulus that is being presented (magnitude estimation), or the numerical values are presented and the participant is asked to adjust the stimulus intensity accordingly (magnitude production). In order to obtain valid estimates of sensations by these direct scaling methods, the estimated or adjusted magnitudes must be meaningful on a ratio scale. Fundamental conditions (axioms) to allow the interpretation of an observer's scaling behavior have been formulated by Narens (1996). Some of these axioms - monotonicity, commutativity, and multiplicativity - are empirically testable. An extension of the axiomatization formulated by Narens (1996) is inherent to Luce's theory of psychophysical scaling, and it allows to fractionate magnitudes of sensations.

According to Narens (1996) and Luce (2002), magnitude estimates or productions are valid only if the commutativity property (or threshold proportion commutativity Luce, 2002) can be proven empirically. Commutativity holds if the outcome of two successive adjustments (e.g., 2× as loud and 3× as loud) is independent of the order in which these adjustments are made. Furthermore, if the adjustments are consistent also with the multiplicative property, then the numbers assigned by the participant may be interpreted as the mathematical numbers they stand for. Multiplicativity (or the probability reduction property Luce, 2002) holds, if two successive adjustments (e.g., 2× as loud and 3× as loud) result in the same stimulus intensity as a single adjustment based on the mathematical product of the two numerical values (e.g., 6× as loud).

These axioms have been tested recently, mostly for the perception of loudness. In a magnitude production experiment, Ellermeier and Faulhammer (2000) showed that whereas commutativity holds (with the exception of one participant), the multiplicativity assumption was violated. That is, doubling the loudness of a 1000-Hz sine tone and then tripling the outcome resulted in the same sound pressure level as first tripling and then doubling the tone. However, in all participants, the resulting sound pressure level exceeded the level which results from a single adjustment to produce a tone that is six times as loud. Similar results have been obtained for the fractionation of loudness (Zimmer, 2005), and in other sensory domains (e.g., in visual size perception; Augustin & Maier, 2008).

The present study is an attempt to test the axioms of commutativity and multiplicativity for the perception of pitch. Perhaps, pitch scaling has been neglected so far because it has been suggested that pitch is not a 'prothetic' sensory continuum characterized by a power function of stimulus intensity (e.g., Stevens & Galanter, 1957). On the other hand, it has been shown that cross-modal matching of pitch with a 'prothetic' continuum (duration) is possible (e.g., Painton, Cullinan, & Mencke, 1977). One problem with the scaling of pitch is that it involves two sensory phenomena, (a) the order of sounds from low to high (tone height) and (b) the similarity between tones that are separated by an octave (tone chroma). To avoid confounds between tone chroma and tone height, participants were asked to fractionate pitch intervals rather than to produce absolute pitch heights in the present study. Additionally, two different starting intervals were compared, one that corresponds to an octave interval, and another one that exceeds an octave interval.

A generalized ratio production procedure (Steingrimsson & Luce, 2005a) was used to obtain fractions of pitch intervals. In this procedure, a lower tone y and a higher tone x is presented as the standard interval, and the participant is asked to match a comparison interval from y to z to a given fraction p of the standard interval. Thus, z can be defined as a tone that makes the subjective interval from y to z stand in the ratio p to the standard interval from y to x (equation 1).

$$x \circ_p y := z \tag{1}$$

With regard to this procedure, the commutativity and the multiplicativity axioms can be formalized as follows:

Commutativity Axiom:
If $x \circ_p y := a$, $a \circ_q y := b$, $x \circ_q y := c$, and $c \circ_p y := d$, then $b = d$

Multiplicativity Axiom:
If $x \circ_p y := a$, $a \circ_q y := b$, and $r = pq$, then $x \circ_r y := b$

Method

Participants

$N = 13$ participants (7 female) were recruited for individual testing in a single-walled sound-attenuated listening room (IAC). Ages ranged between 20 and 55 years ($M = 29.9; SD = 9.6$). All participants had normal hearing with thresholds not exceeding 20 dB HL at any of the audiometric frequencies between 125 and 8000 Hz. Three musically trained participants were included. They reported more than 7 years of musical instruction and continued musical activity.

Stimuli and Apparatus

445 pure sine tones were generated digitally with a sampling rate of 44.1-kHz for all integer frequencies between 259 and 703 Hz. Each tone had a duration of 250 ms including 20-ms cosine-shaped rise and decay ramps. The signals were passed through a Behringer HA 8000 Powerplay PRO-8 headphone amplifier and played back with Beyerdynamics DT 990 PRO (250 Ohm) headphones. The sounds were attenuated to comfortable sound pressure levels of about 70 dB (500 Hz).

Procedure

The experiment consisted of 150 pitch-production trials divided into 3 test sessions. In each session, the participants completed 50 trials, and they were allowed to take a break after 30 trials. In the first session, there was an additional training block consisting of 10 trials which was not included in the analysis. In each trial, two pitch intervals (a lower tone followed by a higher tone) were presented successively to the participant via headphones. The first pair of tones was the standard pitch interval, and the second pair was the comparison pitch interval. The first tones of both intervals were always identical and lower in frequency than the second tones. In accordance with the procedure described by Steingrimsson and Luce (2005b, p. 312), the two tones of an interval were separated by a silent gap of 450 ms, and the two intervals were separated by a 750-ms gap.

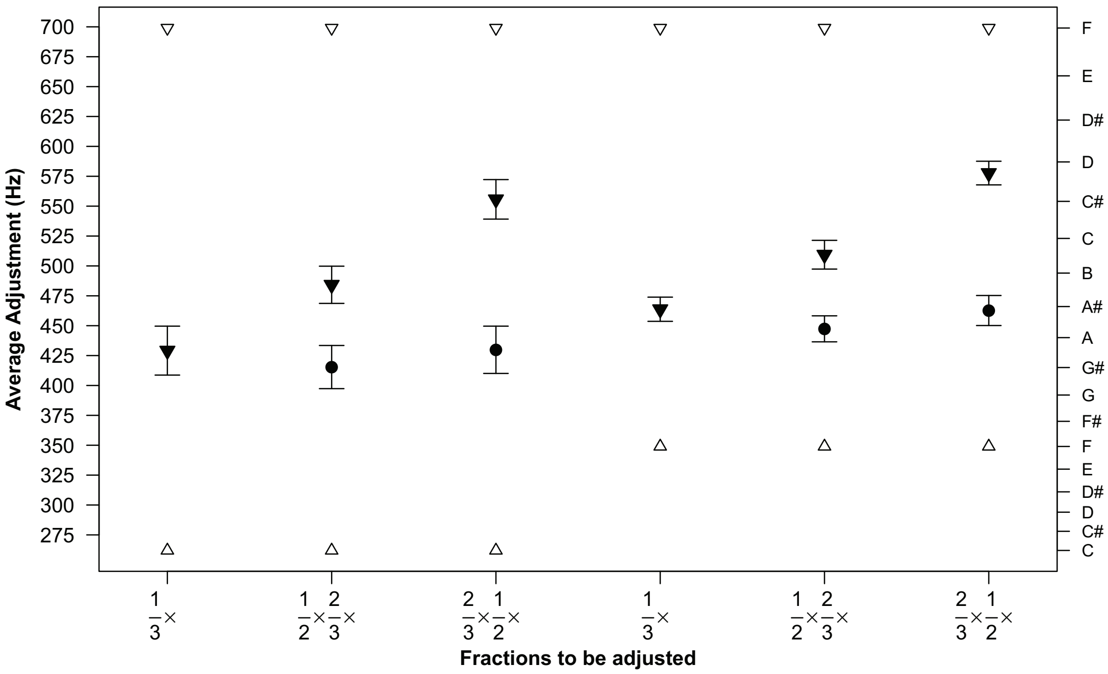
The participants' task was to adjust the size of the comparison interval to a given fraction of the standard pitch interval ($\frac{1}{3}\times$, $\frac{1}{2}\times$ or $\frac{2}{3}\times$) by changing the frequency of the higher tone in the comparison interval. At the beginning of each trial, the to-be-adjusted tone had a randomly chosen frequency lying between the lower tone and 699 Hz. A short instruction together with the respective fraction was displayed on the screen during each trial. The frequency of the higher tone of the comparison interval could be decreased/increased by pressing the left or right cursor key, respectively. The frequency increment/decrement Δf_i was a logarithmic function of the frequency f_i of the tone in that trial: $\Delta f_i = (10 \cdot (log(f_i) - log(262)) + 1)Hz$. By pressing the "Shift" key together with the cursor key, the adjustments could be accelerated (in that case, Δf_i was multiplied by 10). After each adjustment, both pitch intervals were played again (with the last tone having changed in frequency). The participants were encouraged to repeat the adjustments until they were content with the size of the comparison interval. There was no time limit to the task. The adjustments were confirmed by pressing the "Enter" key. The next trial started after a 1-s delay.

Blocks of 10 different types of adjustments were repeated 15 times each. The order of the adjustments within each block was randomized. For two different frequency ranges (starting intervals: 262-699 Hz / C4-F5 and 329-699 Hz / F4-F5), the participants had to produce comparison intervals that were $\frac{1}{3}\times$, $\frac{1}{2}\times$ and $\frac{2}{3}\times$ as large. Additionally, there were two successive adjustments in which either $\frac{2}{3}\times$ or $\frac{1}{2}\times$ of a previously fractionated interval ($\frac{1}{2}\times$ or $\frac{2}{3}\times$ of the starting interval, respectively) had to be adjusted.

Results

The participants made $M = 11.8$ key presses per adjustment on average (ranging between $M_{CL} = 5.4$ and $M_{XS} = 25.3$ individual key presses). There were $M = 5.8 (SD = 3.5)$ accelerated pitch changes, and $M = 5.3 (SD = 3.1)$ small pitch changes per adjustment. In $M = 9.6\%$ (individually ranging between 0% and $M_{JK} = 40.7\%$) of the trials, the participants did not use small-step adjustments. The mean standard deviation of the 15 adjustments per type of trial was 46.2 Hz (ranging between $SD_{MH} = 4.0$ Hz and $SD_{CL} = 77.5$ Hz).

Figure 1. Arithmetic means of individual median adjustments (with standard errors) as a function of the fractionation condition. The open triangles indicate the frequency range; simple fractionations of the starting interval are marked by filled triangles; the results of successive adjustments are depicted by filled circles.



For the total sample, the average median adjustments in the ten fractionation conditions are illustrated in Figure 1. First of all, it can be seen that the instructions to produce one $\frac{1}{3}\times$, $\frac{1}{2}\times$, or $\frac{2}{3}\times$ the starting pitch interval resulted in distinguishable and monotonically increasing adjustments (filled triangles in Fig. 1). Friedman tests revealed that the median $\frac{1}{3}\times$, $\frac{1}{2}\times$, and $\frac{2}{3}\times$ adjustments differed significantly for the large (262-699 Hz), $\chi^2(2) = 24.15; p < .001$, and for the small starting interval (349-699 Hz), $\chi^2(2) = 24.15; p < .001$.

Furthermore, similar comparison intervals were adjusted when the participants had to successively produce $\frac{2}{3}\times$ and $\frac{1}{2}\times$ of the starting interval irrespective of the order (filled circles in Fig. 1). Wilcoxon signed rank tests (two-sided) on the listeners' median adjustments revealed that $\frac{2}{3} \times \frac{1}{2}\times$ and $\frac{1}{2} \times \frac{2}{3}\times$ adjustments did not differ significantly with both starting intervals, $V = 27; p = .21$, and $V = 22; p = .11$, respectively. That is, for the total sample commutativity was not violated.

Additional Wilcoxon signed-rank tests (two-sided) revealed that the participants' median successive adjustments ($\frac{2}{3} \times \frac{1}{2}\times$ and $\frac{1}{2} \times \frac{2}{3}\times$) and their median $\frac{1}{3}\times$ adjustments did not differ significantly both for the large, $V = 44.5; p = .97$, and the small starting interval, $V = 68; p = .12$. This indicates that multiplicativity holds for the total sample, as well.

Violations of the axioms of commutativity and multiplicativity were further tested statistically for the adjustments of each participant. The median individual adjustments together with the significance of axiom violations (two-sided Wilcoxon rank-sum tests) can be seen in Table 1.

Table 1: Median adjustments of the higher tone of the comparison interval (Hz) and evaluation of violations of the axioms of commutativity (p_c) and multiplicativity (p_m) for each listener (monotonicity was significant in all participants). Statistically significant ($p < .1$) axiom violations are printed in boldface. Musically trained participants are indicated by an asterisk.

	Starting interval									
	262-699 Hz					349-699 Hz				
Participant	$\frac{1}{2} \times \frac{2}{3}\times$	$\frac{2}{3} \times \frac{1}{2}\times$	$\frac{1}{3}\times$	p_c	p_m	$\frac{1}{2} \times \frac{2}{3}\times$	$\frac{2}{3} \times \frac{1}{2}\times$	$\frac{1}{3}\times$	p_c	p_m
AK	587	579	578	.31	1	569	572	546	.79	**<.001**
AS*	378	371	372	**.06**	.11	440	448	452	**.01**	**.06**
CL	436	493	549	.46	**.01**	437	513	538	.48	**.04**
CW	426	389	456	.36	**.09**	437	440	450	.16	.11
DD	405	411	379	.63	.87	453	450	477	.69	.10
FK	400	404	376	1	.25	433	430	440	.97	.28
JK	388	408	385	.11	.69	413	426	440	.69	.28
JS	368	398	471	.36	.11	418	473	463	**.08**	1
MH*	360	356	364	**.04**	**.07**	437	441	441	.20	.34
MS	433	422	430	.52	.79	425	456	453	.10	.84
SD	411	435	441	**.01**	.31	447	433	439	.72	.84
WE	483	560	453	**.01**	**.02**	468	518	452	.46	.46
XS*	325	362	325	.21	.32	439	415	438	.11	.80

Discussion

The present investigation shows that instructing listeners to generate different fractions of pitch intervals produces outcomes consistent with the axiom of monotonicity.

There were, however, violations of commutativity, particularly for the large frequency range (4 of 13 participants), though not consistent in magnitude and direction (s. Table 1). This finding is in contrast to what has been reported for other sensory continua (e.g., loudness, Ellermeier & Faulhammer, 2000), and it indicates that the magnitude production of pitch intervals may not be valid on a ratio scale. However, when the frequency range was restricted to an octave, commutativity was violated by only two participants. This suggests that a ratio scale of subjective pitch intervals may exist within but not beyond an octave interval.

Moreover, there were violations of the multiplicativity axiom in some participants. Such violations have also been reported for other sensations (e.g., Augustin & Maier, 2008; Ellermeier & Faulhammer, 2000; Zimmer, 2005), and they indicate that the participants did not use the fractions like scientific numbers. Again, there were more axiom violations for the large starting interval than for the octave starting interval, indicating that tone chroma might interfere with the consistent adjustment of pitch ratios.

References

Augustin, T., & Maier, K. (2008). Empirical evaluation of the axioms of multiplicativity, commutativity, and monotonicity in ratio production of area. *Acta Psychologica*, *129*, 208-216.

Ellermeier, W., & Faulhammer, G. (2000). Empirical evaluation of axioms fundamental to Stevens's ratio-scaling approach: I. Loudness production. *Perception and Psychophysics*, *62*, 1505-1511.

Luce, R. D. (2002). A psychophysical theory of intensity proportions, joint presentations, and matches. *Psychological Review*, *109*, 520-532.

Narens, L. (1996). A theory of ratio magnitude estimation. *Journal of Mathematical Psychology*, *40*, 109-129.

Painton, S. W., Cullinan, W. L., & Mencke, E. O. (1977). Individual pitch functions and pitch-duration cross-dimensional matching. *Perception and Psychophysics*, *21*, 469-476.

Steingrimsson, R., & Luce, R. D. (2005a). Evaluating a model of global psychophysical judgments - I: Behavioral properties of summations and productions. *Journal of Mathematical Psychology*, *49*, 290-307.

Steingrimsson, R., & Luce, R. D. (2005b). Evaluating a model of global psychophysical judgments - II: Behavioral properties linking summations and productions. *Journal of Mathematical Psychology*, *49*, 308-319.

Stevens, S. S. (1956). The direct estimation of sensory magnitudes - loudness. *American Journal of Psychology*, *69*, 1-25.

Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects.* New York: Wiley.

Stevens, S. S., & Galanter, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, *54*, 377-411.

Zimmer, K. (2005). Examining the validity of numerical ratios in loudness fractionation. *Perception and Psychophysics*, *67*, 569-579.

PSYCHOPHYSICAL APPROACHES TO NEURAL EFFICIENCY AND PSYCHOMETRIC INTELLIGENCE: CONTRASTING THE CODING EFFICIENCY HYPOTHESIS AND THE TEMPORAL RESOLUTION POWER HYPOTHESIS.

Thomas H. Rammsayer and Stefan J. Troche
Institute of Psychology, University of Bern, CH-3000 Bern 9, Switzerland
thomas.rammsayer@psy.unibe.ch

Abstract

A central feature of the resource consumption approach to intelligence is that individuals with above-average mental ability utilize the available neural resources more efficiently than below-average individuals. Neural efficiency has also been hypothesized to play a crucial role in perceived duration. According to the coding efficiency hypothesis, perceived duration is considered to reflect the amount of neural activity required for temporal processing of a given stimulus. As an alternative approach, the temporal resolution power (TRP) hypothesis suggests a positive functional relationship between temporal sensitivity and neural efficiency. According to the hypothesis of temporal coding efficiency, mental ability is expected to be negatively related to perceived duration, while the TRP hypothesis predicts a positive correlational relationship between temporal sensitivity and mental ability. The present study was designed to directly test the predictions derived from both these hypotheses. For this purpose, general intelligence (psychometric g) was assessed in 190 participants who also performed an auditory temporal discrimination task. In a next step, individual levels of psychometric g were correlated with psychophysical measures of perceived duration and temporal sensitivity, respectively. While there was a highly reliable positive correlation between temporal sensitivity and psychometric g, no evidence for a functional relationship between psychometric g and perceived duration could be found. The absence of a functional link between perceived duration as an indicator of coding efficiency and psychometric g suggests that coding efficiency reflects aspects of neural efficiency unrelated to psychometric intelligence. On the other hand, temporal sensitivity appears to be a valid psychophysical indicator of aspects of neural efficiency involved in psychometric g.

A central feature of the resource consumption approach to intelligence is that individuals with above-average mental ability utilize the available neural resources more efficiently than below-average individuals. Empirical evidence for this so-called neural efficiency hypothesis of intelligence comes from neuroimaging studies reporting negative correlations between psychometrically measured intelligence and cortical activation as assessed by glucose or oxygen consumption produced by tasks that draw upon these mental abilities (for concise reviews see Neubauer & Fink, 2009; Newman & Just, 2005).

Neural efficiency has also been hypothesized to play a crucial role in perceived duration of intervals in the range of tens to hundreds of milliseconds (Eagleman & Pariyadath, 2009). According to this account, perceived duration is considered to reflect the amount of neural activity required for temporal processing of a given stimulus (Eagleman, 2008; Pariyadath & Eagleman, 2007). Within the framework of the hypothesis of temporal coding efficiency, perceived duration of a stimulus is positively related to neural energy consumption. This means, the more neural energy is necessary for an individual to process the temporal information of a given stimulus, the longer the perceived duration of this stimulus will be. Thus, the physical duration of a presented stimulus should be perceived as longer by an individual characterized by low neural efficiency compared to an individual characterized by high neural efficiency. If an individual's levels of mental ability is positively related to neural efficiency and, at the same time, perceived duration of intervals in the range of milliseconds is negatively related to neural efficiency, it is reasonable to assume a negative functional relationship between mental ability and perceived duration.

An alternative hypothesis suggesting a functional relationship between neural efficiency and temporal information processing is based on the notion of an internal master clock. This concept has been introduced by Surwillo (1968) to account for age-related cognitive impairment and general slowing. He proposed an internal master clock for coordination of different mental activities. More recently, Burle and Bonnet (1999) provided additional converging experimental evidence for the existence of some kind of master clock in the human information processing system. If we assume that the hypothesized internal master clock of individual A works, for example, at half the clock rate as the one of individual B, then A does not only need twice as long as B to perform a specific sequence of mental operations, but also the occurrence probability of interfering incidents will be increased. Thus, A's lower clock rate should lead to worse mental ability compared to B.

Also human timing is often explained by the assumption of a hypothetical internal clock based on neural counting (cf., Rammsayer & Ulrich, 2001). Main features of such an internal-clock device are a pacemaker and an accumulator. The pacemaker emits pulses and the number of pulses relating to a physical time interval is recorded by the accumulator. Thus, the number of pulses counted during a given time interval is the internal representation of the interval. The higher the clock rate, the finer is the temporal resolution of the internal clock which is equivalent to higher temporal sensitivity as indicated by temporal discrimination tasks (for a concise review see Rammsayer & Brandler, 2007).

In a series of experiments, we (e.g., Helmbold, Troche, & Rammsayer, 2007; Rammsayer & Brandler, 2002, 2007) provided converging evidence for the notion that temporal sensitivity as assessed by psychophysical timing tasks may represent a valid indicator of neural efficiency underlying mental ability. More specifically, the Temporal Resolution Power (TRP) hypothesis of intelligence (Rammsayer & Brandler, 2007) proceeds from the assumption that temporal resolution capacity of the brain as assessed with psychophysical timing tasks reflects a major aspect of neural efficiency underlying mental ability. Within this conceptual framework, a positive functional relationship between temporal sensitivity and mental ability is predicted.

The present study was designed to directly test the predictions derived from the hypothesis of temporal coding efficiency and from the TRP hypothesis. According to the hypothesis of temporal coding efficiency, mental ability is expected to be *negatively* related to *perceived duration*, while the TRP hypothesis predicts a *positive* correlational relationship between *temporal sensitivity* and mental ability.

Method

Participants

Participants were 95 male and 95 female volunteers ranging in age from 18 to 39 years (mean±standard deviation of age: 25.0±5.8 years).

Intelligence tests

The aim of psychometric assessment was to obtain a valid measure of general intelligence (psychometric *g*) as an indicator of mental ability. According to Brody (1992), conclusions about psychometric *g* may be unwarranted if they are derived from psychometric intelligence tests limited to a small subset of primary mental abilities. Furthermore, Jensen (1998) emphasized that a composite score will have relatively more psychometric *g* and less specific variance if it is based on a large number of distinct mental tests. Therefore, a comprehensive test battery including 15 subtests was employed for psychometric assessment of different aspects of intelligence. The test battery was composed of several subtests (verbal comprehension, word fluency, space, flexibility of closure, perceptual speed) of the Leistungsprüfsystem (Horn, 1983). In addition, as a measure of performance on reasoning, the short version of the Culture Fair Test Scale 3 (CFT; Cattell, 1961) was applied. Individual CFT scores were obtained on the subscales Series, Classifications, Matrices, and Topologies. Eventually, our test battery comprised two subtests for numerical intelligence and three subtests for verbal, numerical, and spatial memory, respectively, of the Berliner Intelligenzstruktur-Test (Jäger, Süß, & Beauducel, 1997).

Temporal discrimination task

Stimuli were empty auditory intervals marked by a 3-msec onset and a 3-msec offset click. Clicks were white-noise bursts from a computer-controlled sound generator, presented binaurally through headphones at an average intensity of 63 dB SPL. The temporal discrimination task consisted of 64 trials, and each trial consisted of one standard interval and one comparison interval. The duration of the comparison interval varied according to an adaptive rule (Kaernbach, 1991) to estimate x.25 and x.75 of the individual psychometric function; that is, the two comparison intervals at which the response "longer" was given with a probability of .25 and .75, respectively. The standard interval was 50 msec and the initial duration of the comparison interval was 15 msec below and above the standard interval for x.25 and x.75, respectively. In each experimental block, one series of 32 trials converging to x.75 and one series of 32 trials converging to x.25 were presented. Within each series, the order of presentation for the standard interval and the comparison interval was randomized and balanced, with each interval being presented first in 50% of the trials. Trials from both series were randomly interleaved within a block. To initiate a trial, the participant pressed the space bar; auditory presentation began 900 msec later. The two intervals were presented with an interstimulus interval of 900 msec. The participant's task was to decide which of the two intervals was longer and to indicate his/her decision by pressing one of two designated keys on a computer keyboard. After each response, visual feedback ("+", i.e., correct; "-", i.e., false) was provided. As a measure of *temporal sensitivity*, the difference limen (DL) was determined. As a psychophysical indicator of *perceived duration*, the constant error (CE) was computed.

Results

As can be seen from Table 1, performances on the majority of intelligence subtests were significantly correlated with each other. In order to obtain an estimate of psychometric *g*, all psychometric test scores were subjected to a principal components analysis (PCA). The scree criterion was applied to factor extraction. PCA yielded only one factor with an eigenvalue of 4.71 that accounted for 31.4% of total variance. Factor loadings of the intelligence subscales

Table 1. Intercorrelations among intelligence subscales. (V = Verbal Comprehension, W = Word Fluency, S1 = Space 1, S2 = Space 2, C = Flexibility of Closure, P = Perceptual Speed, Ser = Series, Cla = Classifications, Mat = Matrices, Top = Topologies, N1 = Number 1, Number2 = Number 2; vM = Verbal Memory, nM = Numerical Memory, sM = Spatial Memory).

	V	W	S1	S2	C	P	Ser	Cla	Mat	Top	N1	N2	M	nM
W	.54***													
S1	.25**	.30***												
S2	.11	.28***	.42***											
C	.19**	.23***	.38***	.62***										
P	.25**	.35***	.42***	.32***	.31***									
Ser	.17*	.22**	.26**	.37***	.37***	.34***								
Cla	.19*	.21**	.18*	.31***	.28***	.15*	.19**							
Mat	.25**	.34***	.24**	.41***	.44***	.33***	.36***	.27**						
Top	.14*	.21**	.24**	.41***	.42***	.17*	.30***	.21**	.29***					
N1	.49***	.39***	.42***	.29***	.21**	.53***	.29***	.21**	.30***	.25**				
N2	.32***	.28**	.30**	.34***	.37***	.30***	.34***	.25**	.39***	.24**	.45***			
M	.24**	.17*	.13	.02	.12	.23**	.16*	.00	.10	-.07	.15*	.07		
nM	.12	.15*	.21**	.14	.11	.28***	.17*	.13	.15*	.04	.25**	.22***	.34***	
sM	-.00	.13	.23**	.34***	.31***	.21**	.23**	.26***	.26***	.17*	.09	.13	.25*	.29***

* $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed)

on the first unrotated component are shown in Table 2. This first unrotated component is commonly considered an estimate of psychometric *g* (Jensen, 1998).

Mean (±S.E.M.) temporal sensitivity, as indicated by the DL, and mean perceived duration, as indicated by CE, were 18.8±0.64 msec and 2.34±0.41 msec, respectively. While DL revealed that the mean just noticeable difference between a constant 50-msec standard and a variable comparison interval was 18.8 msec, the obtained CE indicated that mean perceived duration of the 50-msec standard interval was 52.34 msec.

Table 2. Factor loadings of intelligence scales on the first unrotated component (psychometric *g*) and communalities (h^2).

Intelligence subscale	Psychometric g	h^2
Verbal comprehension	.50	.25
Word fluency	.58	.34
Space 1	.62	.38
Space 2	.68	.46
Flexibility of closure	.67	.45
Perceptual speed	.64	.40
Series	.59	.34
Classifications	.45	.20
Matrices	.64	.41
Topologies	.50	.25
Number 1	.66	.43
Number 2	.62	.39
Verbal memory	.28	.08
Numerical memory	.39	.15
Spatial memory	.43	.19

Correlational analyses yielded a statistically significant correlation coefficient of *r*=.36 between temporal sensitivity and psychometric *g* (*p*<.001). On the other hand, the correlation coefficient of *r*=-.12 (*p*=.11) between perceived duration and psychometric *g* failed to reach the 5% level of statistical significance. The difference between these two correlation coefficients was highly significant (*z*=3.53, *p*<.001). Because temporal sensitivity and perceived duration were significantly correlated with each other (*r*=-.51, *p*<.001), additional partial correlations were computed. When controlling for perceived duration, the correlation between temporal sensitivity and psychometric *g* remained virtually unchanged (*r*=.35, *p*<.001). Controlling for temporal sensitivity, however, further decreased the non-significant correlation between perceived duration and psychometric *g* from *r*=-.12 to *r*=-.08.

Discussion

Obviously, variation in individual levels of psychometric *g* is linked to aspects of information processing capacity (cf., Jensen, 1998; 2006; Sheppard & Vernon, 2008). To date, however, little is known about the basic biological mechanisms by which this is achieved. Proceeding from the general assumption that psychometric *g* is a function of the central nervous system to process information quickly and correctly, different biological phenomena have been introduced as prime candidates for a biological basis of psychometric *g*: Neuronal refractory periods, reliability of neuronal transmission, neural pruning, myelination of neurons, or differences in neural plasticity. On the whole, all these accounts refer to neural efficiency in the brain as a basic determinant of individual differences in psychometric *g* (Neubauer & Fink, 2009).

These biological approaches to psychometric intelligence can be translated into at least two theoretical accounts related to temporal information processing. According to the coding efficiency hypothesis (Eagleman & Pariyadath, 2009), neural energy expenditure is positively related to perceived duration. More precisely, perceived duration is considered to reflect the amount of neural activity required for temporal processing of a given stimulus (Eagleman, 2008; Pariyadath & Eagleman, 2007). If neural efficiency is involved in both psychometric *g* and coding efficiency, then a *negative* correlational relationship between individual levels of psychometric *g* and perceived duration would be the expected outcome. This is because higher neural efficiency should result in higher psychometric intelligence and shorter perceived duration.

On the other hand, the TRP hypothesis of psychometric intelligence (Rammsayer & Brandler, 2007) predicts finer temporal resolution capacity of the brain as a psychophysical correlate of neural efficiency underlying psychometric *g*. Within this conceptual framework, a *positive* functional relationship between temporal sensitivity and psychometric intelligence could be expected.

Our finding of a non-significant correlation between psychometric *g* and perceived duration clearly argues against the validity of the coding efficiency hypothesis. At the same time, the statistically significant positive relationship between temporal sensitivity and psychometric *g* supports the TRP hypothesis of psychometric intelligence. Given these results, temporal sensitivity appears to be a valid psychophysical indicator of aspects of neural efficiency involved in psychometric *g*. The absence of a functional link between perceived duration, as an indicator of coding efficiency, and psychometric *g* suggests that coding efficiency reflects aspects of neural efficiency unrelated to psychometric intelligence.

References

Brody, N. (1992). *Intelligence*. San Diego, CA: Academic Press.

Burle, B., & Bonnet, M. (1999). What's an internal clock for? From temporal information processing to temporal processing of information. *Behavioural Processes, 45*, 59-72.

Cattell, R. B. (1961). *Culture Free Intelligence Test, Scale 3*. Champaign, IL: Institute for Personality and Ability Testing.

Eagleman, D. M. (2008). Human time perception and its illusions. *Current Opinion in Neurobiology, 18*, 131-136.

Eagleman, D. M., & Pariyadath, V. (2009). Is subjective duration a signature of coding efficiency? *Philosophical Transactions of the Royal Society B, 364*, 1841-1851.

Haier, R. J. (1993). Cerebral glucose metabolism and intelligence. In P. A. Vernon (Ed.), *Biological approaches to the study of human intelligence* (pp. 317-332). Norwood, NJ: Ablex.

Horn, W. (1983). *Leistungsprüfsystem*. Göttingen, Germany: Hogrefe.

Jäger, A. O., Süß, H.-M., & Beauducel, A. (1997). *Berliner Intelligenzstruktur Test Form 4*. Göttingen, Germany: Hogrefe.

Jensen, A. R. (1998). *The g factor*. Westport, CT: Praeger Publishers.

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences*. Amsterdam: Elsevier.

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics, 49*, 227-229.

Helmbold, N., Troche, S., & Rammsayer, T. (2007). Processing of temporal and non-temporal information as predictors of psychometric intelligence: A structural-equation-modelling approach. *Journal of Personality, 75*, 985-1006.

Neubauer, A. C., & Fink, A. (2009). Intelligence and efficiency. *Neuroscience and Biobehavioral Reviews, 33*, 1004-1023.

Newman, S. D., & Just, M. A. (2005). The neural basis of intelligence. In R. J. Sternberg & J. E. Pretz (Eds.), *Cognition and intelligence* (pp. 88-103). Cambridge, UK: Cambridge University Press.

Pariyadath, V., & Eagleman, D. (2007). The effect of predictability on subjective duration. *PLoS ONE, 2*, e1264. (doi:10.1371/journal.pone.0001264)

Rammsayer, T., & Brandler, S. (2002). On the relationship between general fluid intelligence and psychophysical indicators of temporal resolution in the brain. *Journal of Research in Personality, 36*, 507-530.

Rammsayer, T. H., & Brandler, S. (2007). Performance on temporal information processing as an index of general intelligence. *Intelligence, 35*, 123-139.

Rammsayer, T., & Ulrich, R. (2001). Counting models of temporal discrimination. *Psychonomic Bulletin & Review, 8*, 270-277.

Sheppard, L. D., & Vernon, P. A. (2008). Intelligence and speed of information-processing: A review of 50 years of research. *Personality and Individual Differences, 44*, 535-551.

Surwillo, W. W. (1968). Timing of behavior in senescence and the role of the central nervous system. In G. A. Talland (Ed.), *Human aging and behavior* (pp. 1-35). New York: Academic Press.

HOW STABLE IS THE STIMULUS MAGNITUDE EFFECT ON PERCEIVED DURATION?

Martin Verner and Thomas Rammsayer
Institute for Psychology, University of Bern, Switzerland
martin.verner@psy.unibe.ch

Abstract

In the present study, temporal reproduction was used to further elucidate the effect of stimulus magnitude on perceived duration. More specifically, we investigated split-half and test-retest reliabilities of perceived duration as obtained by the temporal reproduction task as well as the effect of stimulus magnitude on perceived duration. For this purpose, 36 participants reproduced the duration of visual stimuli presented for 800, 1000, and 1200 msec, respectively, varying in non-temporal stimulus magnitude. There were two experimental sessions separated by a one-week test-retest interval. A statistically significant main effect of stimulus magnitude on perceived duration could be established with larger stimuli being perceived as longer than smaller ones. As indicated by split-half and test-retest coefficients, perceived durations proved to be highly reliable. Only rather low test-retest coefficients, however, could be demonstrated for the effect of stimulus magnitude on perceived duration indicating poor stability for this latter effect.

Several studies provided empirical evidence for a positive relationship between perceived duration and nontemporal stimulus magnitude. For instance, the perceived duration of visually presented stimuli becomes longer with increasing stimulus magnitude such as size or stimulus complexity (e.g., Long & Beaton, 1980; Ono & Kawahara, 2007, Thomas & Cantor, 1975; Xuan, Zhang, He & Chen, 2007). This effect is referred to as the stimulus magnitude effect on perceived duration. One of the most suitable psychophysical approaches to measure perceived duration represents the so-called temporal reproduction task. Unfortunately, the amount of literature on the reliability of temporal reproduction tasks or the stimulus magnitude effect on perceived duration is very limited. The present study, therefore, was designed to systematically investigate split-half and test-retest reliability of reproduced durations as well as the temporal stability of the stimulus magnitude effect.

Method

Participants

Thirty-six female volunteers, ranging in age from 18 to 28 years (mean age ± SD: 21.3 ± 2.1 years), participated in the experiment. All were undergraduate psychology students from the University of Bern and had normal or corrected-to-normal vision. Informed consent was obtained from each participant prior to the experiment.

Apparatus and Stimuli

The presentation of the stimuli was controlled by E-Prime 2.0 experimental software running on a Dell Optiplex 760 Computer. Participants' responses were logged by means of a Cedrus RB-730 response box. Target stimuli were either filled squares or circles presented in two different sizes subtending a visual angle of 1.2° and 10.0°, respectively. Using a 17" monitor, all stimuli were presented in black color on a white background

Procedure

Participants performed two different visual temporal reproduction tasks a couple of times with a test-retest interval of one week. Order of tasks was randomized across participants. On each task, participants were required to reproduce three different target durations (TD; 800 msec, 1000 msec, and 1200 msec). There were 16 presentations of each TD, resulting in a total of 48 trials per task. Within each trial, the TD was followed by a blank screen for 900 msec (pre-reproduction delay). The start of the reproduction interval was marked by the appearance of a fixation cross. The participants were instructed to end the presentation of the fixation cross by pressing a designated button when its display duration was temporally identical to the corresponding TD. A blank screen between the trials was displayed either for 1000 msec or 1400 msec to prevent a rhythmic response during duration reproduction. To assure the perception of stimulus magnitude information, participants were required to indicate, in addition to the temporal reproduction task, whether the nontemporal target stimulus was a circle or a square (Task A) or whether it was small or large (Task B). More specifically, the participants had to press the left button for duration reproduction (e.g., to terminate the reproduction interval marked by a fixation cross) if the stimulus indicating the TD was small; if a large stimulus was displayed, the right button had to be pressed (see Figure 1). The assignment of response button to hand was held constant within each participant but balanced across participants. Within each task, presentation order of TD, target stimulus magnitude and target stimulus shape was randomized. Prior to each task, instructions were given followed by five practice trials.

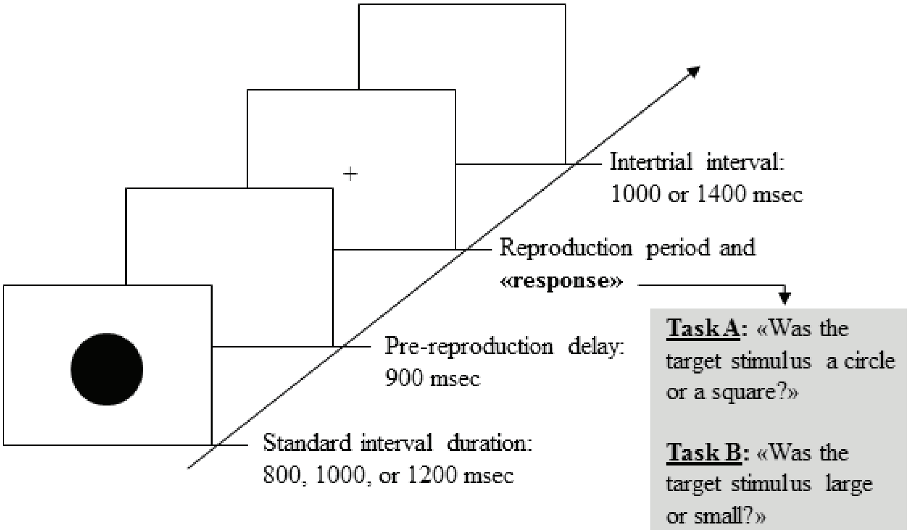


Figure 1. The sequence of events within an experimental trial.

Data Analysis

For each trial, the reproduced duration was logged. The mean reproduced durations (MRD) were compared by means of a three-way analysis of variance (ANOVA) with Magnitude (small and large target stimuli), TD (800 msec, 1000 msec, and 1200 msec) and Time (first and second experimental session) as three repeated-measurement factors. The stimulus magnitude effect was quantified as the difference score between the MRD for large target stimuli and the MRD for small target stimuli. In order to assess the reliability of MRD and of the stimulus magnitude effect, split-half coefficients as derived from the odd-even method were calculated for each experimental session. Resulting coefficients were corrected by means of the Spearman-Brown formula to predict reliability of the whole test length. As to test-retest reliabilities of MRD and the stimulus magnitude effect, coefficients based on the two experimental sessions were calculated by means of Pearson correlations.

Results

Effects on perceived duration

Initial analysis of the data indicated that neither geometrical shape of the target stimuli (squares or circles) nor type of the nontemporal secondary task (Task A and Task B) had an influence on duration reproduction. Hence, for subsequent statistical analyses, MRDs were collapsed across shape of the target stimuli and type of the nontemporal task (Table 1). Greenhouse-Geisser corrected p values are reported where appropriate (for all main effects of TD and the interactions with TD) to protect against violations of sphericity (Geisser & Greenhouse, 1958).

As one would expect, three-way ANOVA yielded a statistically significant main effect TD [$F(1.40, 48.94) = 220.94$; $p < .001$; $\eta_p^2 = .86$] indicating that longer TDs were reproduced longer than shorter TDs. There also was a significant main effect of Magnitude [$F(1, 35) = 34.77$; $p < .001$; $\eta_p^2 = .50$]; large target stimuli were reproduced longer than small target stimuli. Eventually, a significant main effect of Time [$F(1,35) = 21.19$; $p < .001$; $\eta_p^2 = .38$] revealed longer reproduced intervals (i.e., perceived durations in the second experimental session conducted one week after the first session).

All three two-way interactions yielded statistical significance. The interaction between Time and TD [$F(1.80,62.94) = 6.74$; $p < .01$; $\eta_p^2 = .16$] was caused by non-significant differences between MRDs from different TDs (non-orthogonal comparisons). Differences between the two experimental sessions were statistically significant for all three TDs ($p < .001$) as revealed by a post-hoc Scheffé test.

The significant interaction between Magnitude and Time [$F(1,35) = 7.25$; $p < .05$; $\eta_p^2 = .17$] revealed a larger stimulus magnitude effect in the second than in the first experimental session.

Finally, a significant interaction between Magnitude and TD [$F(1.71,59.71) = 8.62$; $p < .01$; $\eta_p^2 = .20$] indicated a more pronounced effect of stimulus magnitude on perceived duration with increasing TD; all orthogonal comparisons yielded statistical significance (Scheffé post hoc test: $p < .01$ for 800 msec; $p < .001$ for 1000 and 1200 msec).

Reliability of reproduced durations and the stimulus magnitude effect

Corrected split-half reliabilities as well as test-retest coefficients for MRDs are presented in Table 2. The split-half reliabilities reached high values, particularly for the second experimental session where coefficients ranged from .94 to .98. The test-retest reliabilities for MRDs were moderate with coefficients ranging from .60 to .71.

Table 1. Mean reproduced durations (MRD) and standard error of mean (SEM) for each target duration, stimulus size, and experimental session.

	Target duration					
	800 msec		1000 msec		1200 msec	
	MRD	SEM	MRD	SEM	MRD	SEM
Measurement 1						
Small stimuli	895	25	1032	27	1148	28
Large stimuli	899	27	1077	31	1223	30
Measurement 2						
Small stimuli	1008	35	1144	37	1226	38
Large stimuli	1077	45	1218	42	1329	46

Table 2. Split-half (odd-even) and test-retest reliability (one-week interval) for mean reproduced durations of the three target durations.

		Target duration		
		800 msec	1000 msec	1200 msec
Measurement 1 (split-half)	Small stimuli	.81	.87	.92
	Large stimuli	.94	.90	.94
Measurement 2 (split-half)	Small stimuli	.98	.94	.96
	Large stimuli	.97	.94	.96
Test-retest reliability	Small stimuli	.70	.65	.60
	Large stimuli	.68	.71	.70

Table 3. Split-half (odd-even) and test-retest reliability for the stimulus magnitude effect obtained with the three target durations.

	Target duration		
	800 msec	1000 msec	1200 msec
Measurement 1 (split-half)	.25	.26	.47
Measurement 2 (split-half)	.43	.71	.72
Test-retest reliability	.10	.01	.13

In contrast to perceived duration, reliability measures of the stimulus magnitude effect were much lower (see Table 3). Despite the reliable MRDs, the stimulus magnitude effect, defined as the difference between MRDs for large and small stimuli, seemed quite unreliable. This especially holds for the first experimental session with split-half coefficients ranging from .25 to .47. Moreover, temporal stability of the stimulus magnitude effect was even worse as indicated by test-retest coefficients not exceeding .13.

Discussion

The present study systematically investigated split-half and test-retest reliability of the measurement of perceived duration and of the stimulus magnitude effect on perceived duration. For this reason, performance on two temporal reproduction tasks was assessed twice in two experimental sessions separated by a one-week test-retest interval. Although, the focus of this work lay on reliability, several effects that could exert some influence on perceived duration will be discussed first.

In both experimental sessions, a large effect of stimulus magnitude on perceived duration was found as indicated by MRDs. Physically larger target stimuli led to longer MRDs compared to smaller target stimuli. There were two types of secondary task employed in the present study. In task A, the participants were instructed to additionally focus on stimulus shape (irrespective of the physical size the target stimulus), whereas in task B, they had to indicate if a presented target stimulus were either physically large or small. The stimulus magnitude effect could be observed independently of the type of the secondary task. These results support the idea of a generalized magnitude system (Walsh, 2003) as well as the notion that magnitude information of a stimulus is processed in an automatic manner (Dehaene & Akhavein, 1995; Schwarz & Ischebeck, 2003).

Moreover, this effect seemed to become more pronounced with increasing TD. To further elucidate to what extent this interaction could be explained by a proportional increase of the stimulus magnitude effect across TDs, MRDs were standardized for TD in an additional analysis. The interaction between Magnitude and TD remained significant ($p < .01$) indicating that the stimulus magnitude effect increases more than just proportionally to TD.

Perceived duration was also influenced by the factor Time. Consistently, the reproduced intervals were longer in the second than in the first experimental session. This result provides additional converging evidence for a repetition effect as described by Hicks and Allen (1979) or Wearden, Pilkington and Carter (1999). These authors explained this effect by declining arousal as consequence of task monotony. Indeed, the tasks used in the present study might have been experienced as rather monotonous and boring. During the second experimental session, this phenomenon might have been particularly strong since participants could anticipate the procedure.

In addition, the significant interaction between Magnitude and Time indicated a more pronounced stimulus magnitude effect in the second compared to the first experimental session. This stronger effect might be explained by a habituation effect of sorts. If one assumes that participants were not familiar with temporal reproduction tasks in the first experimental session, they might have tried out different strategies to memorize and reproduce the presented TDs. In the second experimental session, however, it is reasonable to assume that participants benefited from their experience during the previous session and, thus, applied the strategy that showed the best prove of value during the first experimental session. This notion receives support from our finding of increasing *inter*-individual variability from the first to the second experimental session (see Table 1.) and a concomitant decrease in *intra*-individual variance as reflected by the respective coefficients of variation (standard deviation of reproduced duration divided by MRD).

For MRD, the split-half coefficients ranging from .81 to .98 revealed high reliability within an experimental session. Although being lower than split-half reliabilities, test-retest coefficients ranging from .60 to .70 suggested MRDs to be moderately stable across the one-week test-retest interval. Neither split-half coefficients nor test-retest coefficients appeared to be differentially affected by the type of the secondary task.

The stimulus magnitude effect on perceived duration - quantified as the difference between MRDs for large and small stimuli – was shown to be less reliable than the MRDs. Split-half reliabilities were rather inconsistent and ranged from .25 to .72 while the stimulus magnitude effect seemed to be more consistent in the second compared to the first experimental session. Such a difference might be explained by the participants' higher precision in reproduced durations (lower intra-individual variance) in the second experimental session. However, with test-retest coefficients between .01 and .13, the stimulus magnitude effect proved not to be stable across time. A possible explanation for this obvious lack of temporal stability could be seen in the statistical method applied. It is well known, that the reliability of difference scores is rather low, especially if underlying variables are highly correlated to each other (Cohen & Cohen, 1983). Because the MRDs for large and small stimuli were highly correlated in the present study, the difference scores may include a large amount of measurement error. Thus, the stability of the magnitude effect on perceived duration could be somewhat underestimated. Nevertheless, split-half coefficients in the second experimental session reached moderate values for the 1000- and 1200-msec TDs, respectively, despite the use of difference scores. Thus, the large discrepancy between split-half and test-retest coefficients could be additional support for a weak stability of the stimulus magnitude effect across time. Furthermore, the available data do not support the idea of certain individuals being more sensitive to stimulus magnitude compared to others.

In conclusion, the tasks employed in the present study proved to be well suited to investigate the stimulus magnitude effect on perceived duration. The mean reproduced durations were highly reliable within one testing session and moderately stable across time. However, even if partially consistent within a given experimental session the effect of stimulus magnitudes on perceived duration does not seem to be stable across time.

References

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Dehaene, S., & Akhavein, R. (1995). Attention, automaticity, and levels of representation in number processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 314-326.

Geisser, S., & Greenhouse, S.W. (1958). An extension of Box's results on the use of the *F* distribution in multivariate analysis. *Annals of Mathematical Statistics 29*, 885-891.

Hicks, R., & Allen, D. (1979). The repletion effect in judgments of temporal duration across minutes, days, and months. *American Journal of Psychology, 92*, 323-333.

Long, G. M., & Beaton, R. J. (1980). The contribution of visual persistence to the perceived duration of brief targets. *Perception & Psychophysics, 28*, 422-430.

Ono, F., & Kawahara J.-I. (2007). The subjective size of visual stimuli affects the perceived duration of their presentation. *Perception & Psychophysics, 69*, 952-957.

Schwarz, W., & Ischebeck, A. (2003). On the relative speed account of number-size interference in comparative judgments of numerals. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 507-522.

Thomas, E. A. C., & Cantor, N. E. (1975). On the duality of simultaneous time and size perception. *Perception & Psychophysics, 18*, 44-48.

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences, 7*, 483-488.

Wearden, J. H., Pilkington, R., & Carter, E. (1999). "Subjective lengthening" during repeated testing of a simple temporal discrimination. *Behavioral Processes, 46*, 25-38.

Xuan, B., Zhang, D., He, S., & Chen, X. (2007). Larger stimuli are judged to last longer. *Journal of Vision, 7*, 1-5.

ASSIMILATION BETWEEN TWO NEIGHBORING TIME INTERVALS MARKED WITH TACTILE STIMULI

Emi Hasuo, Tsuyoshi Kuroda, Simon Grondin
Université Laval, Québec, Canada
emi.hasuo.1@ulaval.ca

Abstract

The aim of this study was to examine whether and how the perception of an empty time interval marked by two brief tactile stimuli, S, is influenced by the duration of the preceding time interval, P. In each trial, three electric pulses, each lasting 20 ms, were successively delivered to the right hand of participants, the first two pulses marking P and the last two marking S. S was fixed at 240 ms while P was 160, 240, or 320 ms. In addition, there was a control condition where S was presented in isolation (without P). Perceived duration of S was estimated with the method of constant stimuli. Results showed that participants underestimated S when P was 160 ms. This underestimation appeared as a kind of perceptual assimilation between P and S, but indeed S was not overestimated when it was preceded by a longer interval (P = 320 ms). The underestimation was rather regarded as the time-shrinking illusion which had been tested with visual and auditory stimuli.

One of the main issues in time perception studies has been whether or not the brain has a single dedicated time-keeping device, called a *central* clock, which processes duration in general irrespective of the sensory modality of inputs (e.g., Grondin, 2010). This device is assumed to consist of two modules, one emitting neural pulses which are accumulated in the other; the amount of accumulated pulses determines the perceived duration of intervals. The beginning and end of accumulation are triggered by sensory inputs while the accumulation process, per se, is not affected by the property of sensory inputs (e.g. Gibbon et al., 1984). However, this notion is being challenged by some empirical findings. For example, Grondin and Rousseau (1991) indicated that sensitivity to time differs between sensory modalities. In their experiment, intervals were better discriminated (the Weber fraction was lower) when they were marked by auditory stimuli than when marked by visual or tactile ones, and the discrimination was even much lower when an empty interval was marked by two brief signals delivered from different sensory modalities. This finding is inconsistent with that expected from the internal-clock hypothesis; if inputs from different modalities were processed by a single dedicated device, discrimination levels should be relatively constant regardless of marker modalities.

On the other hand, the internal-clock hypothesis would receive empirical support would one find a common time-perception phenomenon occurring in different sensory modalities. In other words, a phenomenon should occur in more than one modality if its occurrence is attributed to the central-clock mechanism that is independent from stimulus modalities. Indeed, there is an illusion that appears in both visual and auditory modalities. This illusion, called *time shrinking*, has been first reported by Nakajima and ten Hoopen (1988), and it has been investigated mainly in the auditory modality (see ten Hoopen et al., 2008 for a review). When three successive short sounds mark two neighboring time intervals (t_1 and t_2), the subjective duration of the second interval (t_2) can be shortened by a considerable degree under certain time conditions: when t_1 is shorter than t_2, and this difference is smaller than about 100 ms. Time shrinking appears maximally for short durations, i.e., when $t_1 < 200$ ms. Since in most cases the subjectively shortened t_2 approaches the subjective duration of t_1, time shrinking can be considered to be a type of temporal assimilation between two neighbouring time intervals (Nakajima et al., 2004). A later study by Arao et al. (2000) showed that time shrinking can appear also when the time intervals are marked by short flashes (although the time condition in which time shrinking takes place differed: it took place for a wider range of time conditions in vision compared to audition). Their results indicated that the visual system, like the auditory one, can assimilate the second interval to the first one. The fact that time shrinking appears for both auditory and visual stimuli suggest that the processing of temporal information in the auditory and the visual system relies on a single device.

Given that several perceptual phenomena take place in more than one modality (e.g., the *tau* and *kappa* effects; see Jones & Huang, 1982), it seemed reasonable to predict that time shrinking can occur in other sensory modalities. In the present study, we investigated whether or not time shrinking takes place in tactile modality. Since this was the first time that time shrinking was examined with tactile stimuli, we included an experimental condition which would cause time shrinking in both audition and vision, i.e., $t_1 = 160$ and $t_2 = 240$ ms. If it is shown that time shrinking occurs also for tactile stimuli, it will suggest that the timing system for auditory, visual, and tactile modalities share a single device.

Method

Participants

Twelve participants (9 females and 3 males) were recruited at Laval University. The mean age was 27.8 years (*SD* = 7.02). Each participant was paid CAD$28.

Stimuli and Apparatus

Each tactile stimulus was a 20-ms electric signal, delivered to the right hand of the participant. The intensity of the stimuli was adjusted to a comfortable level for each participant. The stimulus patterns are illustrated in Figure 1.

In the Experimental condition, the standard pattern consisted of three successive stimuli, the first and the second marking the preceding time interval (P), and the second and the third marking the standard time interval (S). In the control condition, the standard pattern consisted of two successive stimuli marking S. In both cases, the comparison interval (C) was marked by two successive stimuli.

S was fixed at 240 ms, and P was 160, 240, or 320 ms. The condition /160/240/ (slashes denote short markers delimiting P and S, as /P/S/) was the temporal pattern in which considerable underestimation occurred in both the auditory and the visual modality (Nakajima et al., 2004; Arao et al., 2000). The number of conditions was 4 (3 P-durations + 1 control). The duration of C was 100, 140, 180, 220, 260, 300, 340, or 380 ms. A computer application (E-prime) was used to generate the stimuli and to make a program steering the course of the experiment. The stimuli were presented from electrodes fixed to the back of the participant's right hand, one between the index finger and the middle finger, and the other between the ring finger and the pinky. The electrodes were connected to a computer (IBM Netvista) via an amplifier (Solutions Temps Reel).

Procedure

The method of constant stimuli was used. Participants were instructed to judge whether C was subjectively shorter or longer than S, and to respond "shorter," "longer," or "unsure" by pressing a button with their left hand (we instructed the participants to try to respond "shorter" or

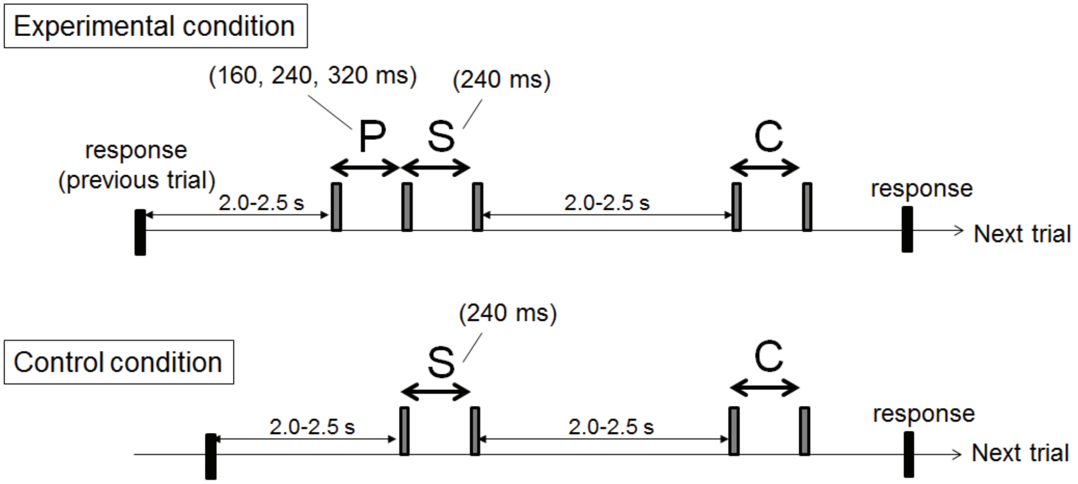


Figure 1. Stimulus presentation chart for the experiment. The durations of P, S, and C correspond to the temporal distance between the beginning of one marker and the beginning of the subsequent marker.

"longer" and to try not to respond "unsure" when possible). For each condition, the 8 durations of C (100-380 ms) were presented in random order.

Each condition was blocked together to be conducted in one of four experimental sessions: one for /160/240/, one for /240/240/, one for /320/240/, and one for the control condition. There were 240 trials per session (8 C-durations × 30 repetitions), and these trials were divided in 3 blocks of 80 trials. Participants were able to take a break between blocks.

The order of the three experimental conditions was counterbalanced (there were 6 possible orders). Half of the participants did the three experimental conditions first and then moved on to the control condition, while the other half of the participants started with the control condition and then moved on to the three experimental conditions. Thus, there were 12 possible orders: 6 for the experimental condition × 2 for the order of experimental and control.

Results

There were 11,520 responses in total (4 conditions × 8 C-durations × 30 repetitions × 12 participants). The number of "unsure" responses was 359, and they were split evenly between "longer" and "shorter" responses. For each condition and each participant, a psychometric function was obtained by plotting the probability of responding that C was "longer" against the C-duration. Figure 2 shows the psychometric functions using the "longer" probabilities averaged across participants.

The cumulative normal distribution was fitted to the psychometric function. The duration of C which corresponded to a probability of 0.5 for responding that C was "longer" was considered as the *point of subjective equality* (*PSE*) of S. The standard deviation was also calculated for each curve. A lower standard deviation means a steeper slope of the curve, and thus indicates better discrimination (see Grondin, 2008). Thus, the PSE and the standard

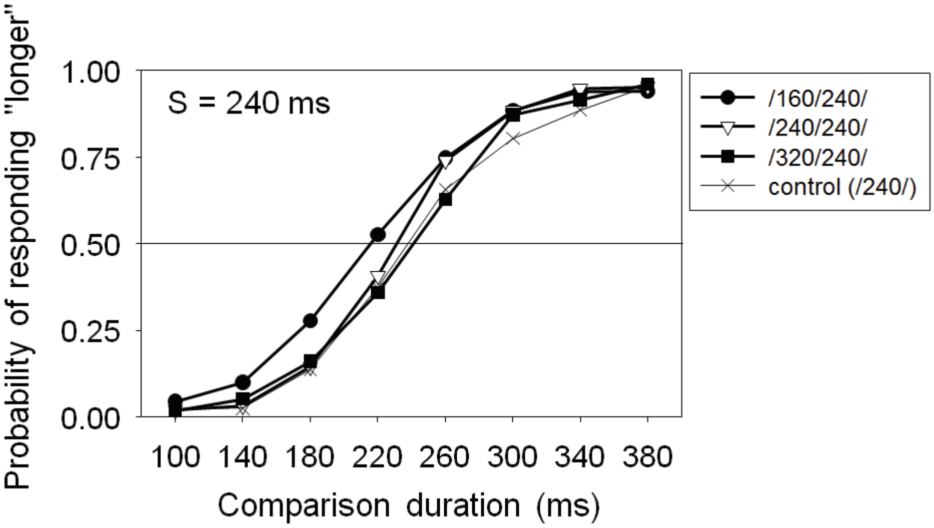


Figure 2. Probability of responding that the comparison interval was longer than the standard (240 ms) in each condition. Each data point is based on 360 responses (30 responses by each of the 12 participants).

deviation were obtained for each condition for each participant. The PSE was smallest when P = 160 ms, and increased as P became longer.

To eliminate experimental bias and to compare the effect of the P durations, we calculated the amount of overestimation of S for the three experimental conditions by subtracting the PSE for the control condition from the PSE for the experimental conditions (i.e. PSEexp – PSEcont) for each participant.

Figure 3 shows the mean overestimation of the 12 participants. The overestimation was negative (i.e., S was underestimated) when P = 160 ms, and the underestimation decreased as P lengthened.

A one-way repeated-measures ANOVA with P-duration as factor was performed using the amounts of overestimation. The effect of the P-duration was significant, $F(2, 22) = 6.649$, $p < .01$, $\eta_p^2 = .377$. Bonferroni post-hoc test revealed that the difference between P = 160 ms and P = 240 ms, and between P = 160 ms and P = 320, ms were significant ($p < .05$).

Discussion

The results of the experiment clearly showed that, when two time intervals neighbor each other, the subjective duration of the second interval can be influenced by the first one. When the first time interval was equal to or shorter than the second, the duration of the second time interval was underestimated (Figure 3).

The underestimation of S was in line with the previous studies on time shrinking; time shrinking in the auditory modality was said to take place typically when (1) $0 \leq S - P \leq 100$ ms, and (2) $P < 200$ ms (e.g. ten Hoopen et al., 2008). In our experiment, the time condition in which maximum time shrinking took place, /160/240/, was within this range (e.g., Nakajima et al., 2004). For this condition, time shrinking had been reported also in the visual modality (Arao et al., 2000). We hereby demonstrated that time shrinking occurs also in the tactile modality for the same time condition.

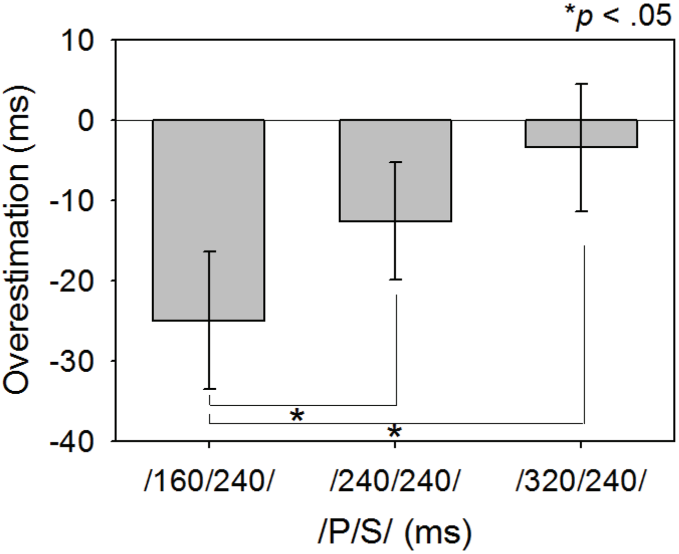


Figure 3. Mean overestimation of S in each experimental condition (PSEexp – PSEcont). Negative overestimation indicates time shrinking. Error bars show the standard error of mean (SEM). Results of multiple comparisons based on Bonferroni's method are also shown.

The underestimation of the S when /160/240/ must have made the subjective duration of S approach the subjective duration of the shorter P. In other words, this underestimation could be understood as perceptual *assimilation*.

It was interesting that S was underestimated even when P and S had the same physical duration (see the /240/240/ condition in Figure 3). Although underestimation of S may seem unnecessary in terms of temporal assimilation when P = S, underestimation of S in such conditions had been shown repeatedly in the figures of previous auditory and visual experiments, especially when the preceding interval was short (e.g., Arao et al., 2000; Miyauchi & Nakajima, 2005; Nakajima et al., 2004). It could be possible that, when P = S, P is already slightly underestimated, so the following S must be shortened to approach the subjective duration of P (Miyauchi & Nakajima, 2005).

It should be noted that, when P was longer than S, perceptual assimilation did not occur: if assimilation had taken place, the S should have been overestimated to approach the subjective duration of the longer P. However, such overestimation of S did not appear (see the /320/240/ condition in Figure 3). This is interesting, for in both /160/240/ (in which assimilation occurred) and /320/240/ (in which assimilation did not occur), the difference between the duration of P and S was 80 ms. It seems that the range in which temporal assimilation occurs is asymmetric, as in the auditory modality (-80 ms ≤ [P – S] ≤ 40 ms; ten Hoopen et al., 2008).

In summary, we found that time shrinking, an illusion in temporal perception that had been reported mainly in auditory modality, takes place also with tactile stimuli. The occurrence of time shrinking (temporal assimilation) was similar to that reported previously for auditory and visual modalities, which may indicate that the timing system for auditory, visual, and tactile modalities share a similar process to some extent. It should be interesting to investigate the occurrence of time shrinking further with a wider range of time intervals, which could allow more thorough comparison between modalities.

Acknowledgement

This research was made possible by a research grant awarded to SG by the Natural Sciences and Engineering Council of Canada (NSERC). We would like to extend special thanks to Félix Désautels, Katherine Labonté, and Vincent Laflamme for their help in data collection.

References

Arao, H., Suetomi, D., & Nakajima, Y. (2000). Does time-shrinking take place in visual temporal patterns? *Perception*, *29*, 819-830.

Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. In J. Gibbon & L. Allan (Eds.), *Timing and time perception* (Annals of the New York Academy of Sciences, Vol. 423, pp. 52-77). New York: New York Academy of Sciences.

Grondin, S. (2008). Methods for studying psychological time. In S. Grondin (Ed.). *Psychology of time* (pp. 51-74). Bingley, UK: Emerald Group Publishing.

Grondin, S. (2010).Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception & Psychophysics, 72*, 561-582.

Grondin, S., & Rousseau, R. (1991). Judging the relative duration of multimodal short empty time intervals. *Perception & Psychophysics*, *49*, 245-256.

Jones, B., & Huang, Y. L. (1982). Space–time dependencies in psychophysical judgment of extent and duration: Algebraic models of the tau and kappa effects. *Psychological Bulletin*, *91*, 128-142.

Miyauchi, R., & Nakajima, Y. (2005). Bilateral assimilation of two neighboring empty time intervals. *Music Perception*, *22*, 411-424.

Nakajima, Y., & ten Hoopen, G. (1988). The effect of preceding time intervals on duration perception. *Proceedings of the Acoustical Society of Japan*, Autumn meeting (pp. 381-382).

Nakajima, Y., ten Hoopen, G., Sasaki, T., Yamamoto, K., Simons, M., & Suetomi, D. (2004). Time-shrinking: The process of unilateral temporal assimilation. *Perception*, *33*, 1061-1079.

ten Hoopen, G., Miyauchi, R., & Nakajima, Y. (2008). Time-based illusions in the auditory mode. In S. Grondin (Ed.). *Psychology of time* (pp. 139-188). Bingley, UK: Emerald Group Publishing.

MEASUREMENTS OF VELVET HAND ILLUSION BY MAGNITUDE ESTIMATION AND PAIRED COMPARISON

Tetsu Miyaoka
Faculty of Comprehensive Informatics, Shizuoka Institute of Science and Technology,
2200-2 Toyosawa, Fukuroi 437-8555, Japan
Miyaoka@cs.sist.ac.jp

Abstract

The purpose of this study was to measure the level of the Velvet Hand Illusion (VHI) and to propose mechanisms that produce the VHI. The experimental stimuli were instruments with two straight rods positioned parallel to each other. The participant held the two rods between his/her hands and moved both hands simultaneously in the direction orthogonal to the rods. The participants judged the level of the VHI by the methods of magnitude estimation and paired comparison. Both methods gave the highest level of the VHI when the distance between the rods was 100 mm, and the lower level of the VHI when the distance was larger or smaller than 100 mm. In this paper, we described the experiments and presented a simulation model to explain the VHI levels, and we suggested the mechanisms that produce the VHI.

One can easily feel the Velvet Hand Illusion (VHI) by holding a coarse-wire net between both hands and moving them simultaneously on the net. The surface of the contralateral hand feels very soft and smooth, as if one is touching the surface of velvet. But, wire is not needed for experiencing the VHI: two rods positioned parallel to each other (shown in Fig. 1) will also suffice. If one holds the two rods between his/her hands and moves both hands simultaneously in the direction orthogonal to the rods, the same perception as that by the coarse-wire net can be experienced.

The VHI is one of the clearest illusions in tactile sensation. But only a few studies have been conducted and the mechanisms to produce the VHI have not been found (Ohka et al., 2010; Rajaei et al., 2012). The purpose of the current study was to measure the level of the VHI by two methods, magnitude estimation and paired comparison. We produced a mathematical model to explain the level of the VHI, and we proposed the brain mechanisms that make the VHI possible.

Experiment 1

The purpose of Experiment 1 was to determine the level of the VHI. In this experiment, we adopted the magnitude-estimation method, which uses a modulus.

Method

Participants: Eight males and one female in their twenties participated in Experiment 1. None of the participants had prior experience of this type of experiment.
Stimuli: Two straight rods were used as stimuli. As shown in Figure 1, the rods were set in a U-shaped frame. The two rods were positioned parallel to each other. The diameters of the rods and the frame were 3 mm and 8 mm, respectively. The rods and the frame were made of

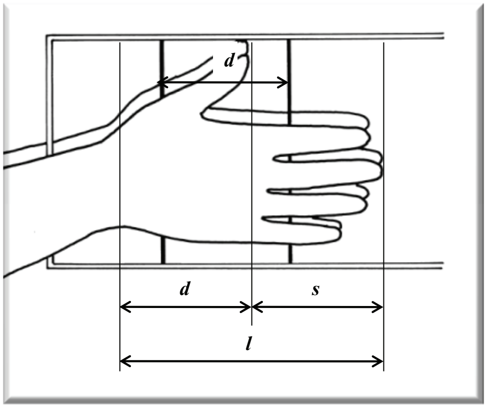


Figure 1. An example of the stimuli used in Experiments 1 and 2. The participant held two rods between both hands and moved the hands simultaneously in the orthogonal direction to the rods. The distance between the two rods is *d*. The length of the hand movement is *s*. The length of the palm region is *l*.

iron covered by plastic. Seven distances between the rods (*d* in Figure 1) were used in Experiment 1: 20, 40, 60, 80, 100, 120, and 140 mm.

Procedure: The participant was seated in a chair and wore an eye mask to prevent visual inspection of the stimuli. Two stimuli, one that was a standard stimulus and the other one that was a comparison stimulus, were set on a desk in front of the participant. The two stimuli were set on the right and left sides from the participant's viewpoint and 25 cm apart from each other. In the magnitude-estimation procedure, a modulus was used. The stimulus with the 80 mm distance between rods was chosen as the "standard" stimulus and given the modulus number “100”. The participant was asked to assign a number as the illusion level of the comparison stimulus when it was compared with the illusion level of the standard stimulus. The experimenter informed the participant which was the standard stimulus in the two stimuli set on the right and left sides.

The maximum presentation time of one set of stimuli was 15 seconds and inter-stimulus interval was 20 seconds. Each participant responded 20 times for each stimulus set. Therefore, the total number of experimental trials for each participant was 140. The presentation order of each set was random.

The temperature in the laboratory was maintained at higher than 25ºC in the experimental period to avoid the decrease of tactile sensation.

Results and discussion

The geometric mean values of the magnitude estimation were calculated based on the experimental data of the nine participants. The results are shown in Figure 2. The diamonds in the figure are the means of the magnitude estimations.

This figure shows that the mean values of the magnitude estimations increased when the distances between the two rods increased from 20 mm to 80 mm, and reached a maximum at 80 mm (mean: 102) and at 100 mm (mean: 103). The magnitude-estimation values decreased when the distances increased to more than 100 mm. For example, the magnitude-estimation values were 66.1 at 140 mm.

The participants reported that the tactile impression of the two rods was strong in the

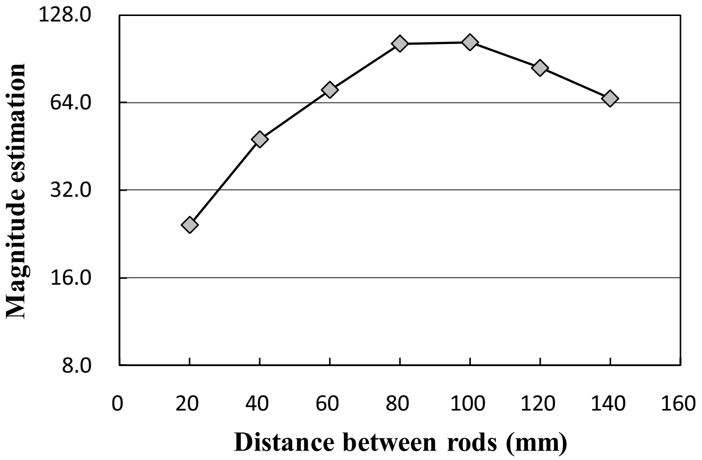


Figure 2. Levels of the VHI measured by the magnitude-estimation method. The vertical axis shows the magnitude-estimation value and the horizontal axis shows the distance between the rods. Each diamond symbol in the figure shows the geometric mean of the magnitude-estimation value.

20 mm distance condition. The shorter distance masked the illusion and the VHI became weak. In contrast, the participants reported that the 140 mm distance was too large to allow enough hand movement (*s* in Figure 1) and the illusion became weaker.

The results of the magnitude-estimation experiment showed that the levels of the VHI were determined by the distances between the rods. The curve of the illusion level was convex upward and the VHI were at the highest level when the distances of the rods were between 80 and 100 mm. The illusion level decreased below 80 mm and above 100 mm, but the causes for making the illusion decrease might be different between the short and long distances of the rods.

Experiment 2

The purpose of Experiment 2 was to determine the levels of the VHI by the method of paired comparison. The relative magnitudes of VHI were calculated using Thurstone's Case V method.

Method

Participants: Ten males and one female in their twenties participated in Experiment 2. None had prior experience of a paired-comparison experiment.
Stimuli: The same types of stimuli as in Experiment 1 (shown in Figure 1) were adopted in Experiment 2. Seven stimuli were given with the following distances between rods (*d* in Figure 1): 80, 85, 90, 95, 100, 105, and 110 mm.
Procedure: The participant was seated in a chair and wore an eye mask. Two stimuli were set on the desk. The method for the stimulus setting was the same as that for Experiment 1. The participant touched the two stimuli successively and compared the levels of the VHI of the two stimuli. The participant determined which stimulus produced a larger illusion and responded by the two-alternative forced-choice technique.

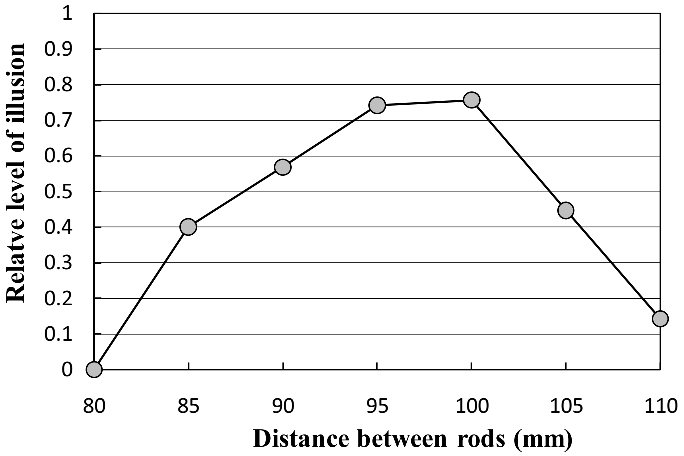


Figure 3. The VHI measured by the method of paired comparison. The vertical axis shows the relative level of illusion. The horizontal axis shows the distance between the rods. Each symbol in the figure shows the relative level of the VHI.

The number of combinations for choosing two stimuli from the seven distinct stimuli was 21. Since it was necessary to counterbalance the right and left side positions of the stimuli, the total number of combinations was 42. The maximum presenting time of a stimulus set was 15 seconds and the inter-stimulus interval was 20 seconds. Each participant performed 10 trials for each combination. Therefore, the total number of experimental trials was 420 for each participant.

Other experimental conditions were the same as those in Experiment 1.

Results and discussion

The relative VHI levels were calculated by Thurstone's Case V method, based on the paired-comparison data of the eleven participants. The results are shown in Figure 3. The circles in the figure show the relative levels of the VHI. The illusion level was set to zero when the distance between the two rods was 80 mm. The illusion levels increased from 80 mm, reached the maximum at 100 mm (relative level of illusion = 0.757), and decreased above 100 mm.

The patterns of the curves of the magnitude estimation (Experiment 1) and of the paired comparison were generally similar to each other. However, their details were different. The illusion levels in the magnitude-estimation experiment did not change much between 80 mm and 120 mm (80 mm: mean 102, 100 mm: mean 103, and 120 mm: mean 84.2). However the illusion levels in the paired-comparison method greatly changed; the illusion levels in the paired-comparison method were a relative one and, generally speaking, the paired-comparison method is more accurate than the magnitude-estimation method.

From the results of both experiments, we induced the following hypothesis:

(1) When the distance between two rods is short, the impression of the rods is strong and masks the VHI. By increasing the distance, the levels of the VHI become high.

(2) The VHI level is high when the distance between the two rods is not too long and the participant can move his/her hands enough to touch both rods. The illusion levels decrease if the participant cannot touch both rods simultaneously when he/she moves both hands.

General discussion

We explain the simulation model of the VHI first and then propose the mechanisms that make the VHI possible.

Simulation model of the VHI

Suppose that the level of the VHI is determined by the length of the palm region (l in Figure 1), the distance between the two rods (d in Figure 1), and the length of the hand movement (s in Figure 1). The participant holds the rods between his/her hands and sets the wrist joint at the same position of the proximal side of the rod and moves both hands in the proximal direction.

The VHI occurs at a maximum level when $0 < d \le l - sc$, where sc is a constant to determine the beginning point of the VHI decrease. The VHI decreases when $l - sc < d \le l$. The VHI does not occur when $d > l$, because the participant touches only one rod. These conditions are modeled as

$$\begin{cases} f_1(d) = 1 & (0 < d \le l - sc) \\ f_1(d) = -\dfrac{1}{sc^2}\{d - (l - sc)\}^2 + 1 & (l - sc < d \le l) \end{cases} \tag{1}$$

Equation (1) is shown by the diamonds and the dotted line in Figure 4, where l = 170 mm and sc = 90 mm.

The level of the VHI also changes depending on the distance between the two rods. When the distance is short, the impression of the rods is strong and masks the illusion. When the distance becomes longer, the VHI becomes larger. These relations are modeled as

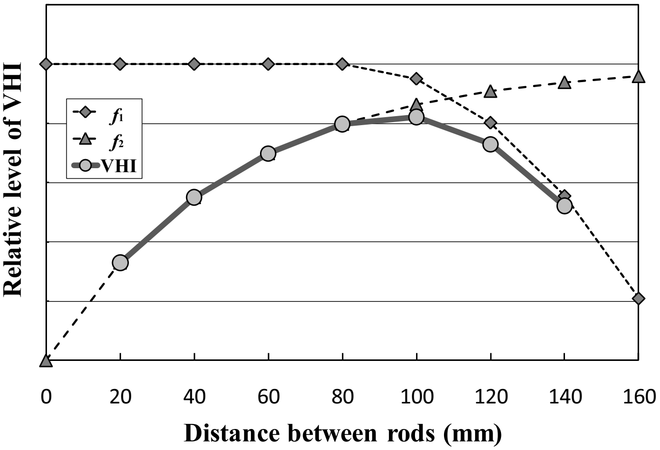


Figure 4. The simulation model of the VHI. The terms f_1 and the f_2 in the legend mean $f_1(d)$ and $f_2(d)$, respectively. The levels of the VHI are determined by multiplication of $f_1(d)$ and $f_2(d)$ and are shown by the thick line.

$$f_2(d) = \frac{e^{ad} - 1}{e^{ad}} \qquad (2)$$

where a is a time-constant parameter. Equation (2) is shown by the triangles and the dashed line, where $a = 0.02$.

The total levels of the VHI are presented as the multiplication of f_1 (d) and f_2 (d) and are shown by the circles and the thick line in Figure 4. The simulation model explains, in particular, the convex curve of Experiment 1 well.

Inference of mechanisms that make the VHI possible

We proposed the emergence mechanisms of the VHI as follows.

The participant holds a coarse-wire net between his/her hands and moves both hands simultaneously. He/she touches the wire net and the surface of his/her hands. The wire net moves against the hands, but the hands do not move against each other because the participant moves both hands simultaneously. The contralateral surface of the hand gives no friction. The brain of the participant infers that the hands are moving on the surface of some objects. But the contralateral surface of the other hand gives no friction. Therefore, the brain concludes that the contralateral surface of the hand is very smooth and soft.

Conclusion

In this study, we performed magnitude-estimation and paired-comparison experiments, and showed that the level of the Velvet Hand Illusion was the highest when the distance of the rods was 100 mm. We presented a simulation model to explain the level of the VHI, and proposed mechanisms that produce the VHI. Further studies are needed to reveal the neural mechanisms that produce the VHI.

References

Ohka, M., Kawabe, Y., Abdullah, C., Rajaei, N., Yussof, H. B., and Miyaoka, T. (2010). Investigation on Velvet Hand Illusion using psychophysics and FEM Analysis. *International Journal of Smart Sensing and Intelligent Systems*, 3, 488-503.

Rajaei, N., Kawabe, Y., Ohka, M., Miyaoka, T., Chami, A., and Yussof, H. B. (2012). Psychophysical experiments on Velvet Hand Illusion toward presenting virtual feeling of material. *International Journal of Social Robotics*, (online).

Effect of space between fingers on tactile duration discrimination

Tsuyoshi Kuroda, Simon Grondin
Université Laval, Québec, Canada
tsuyoshi.kuroda.1@ulaval.ca

Abstract

The purpose of this study was to examine how space between tactile stimuli marking time affects the processing of duration. Twelve participants were recruited in an experiment where the method of constant stimuli was used. Eight participants perceived empty time intervals as longer when the intervals were marked by two electric pulses stimulating different (middle and little) fingers than the same (middle) finger while four participants perceived intervals as shorter when two pulses were delivered to different fingers. However, increasing the angle between the middle and little fingers did not magnify the spatial effect in both groups of participants. Finally, discrimination level remained the same regardless of whether participants overestimated or underestimated intervals marked by two pulses stimulating different fingers.

Grondin, Kuroda and Mitsudo (2011) examined how duration processing of inter-stimulus intervals is changed when the intervals are marked by two electric pulses stimulating different hands instead of the same hand. They reported that delivering two pulses to different hands led to longer perceived duration than delivering them to the same hand. They argued that this effect is a variation of the *kappa effect*, where an empty time interval is perceived as longer when it is bounded by two signals further away from each other in space. The kappa effect has been tested in the visual and auditory modalities, and even in the tactile modality (see Suto, 1952, 1955; Yoblick & Salvendy, 1970). Indeed, in Grondin et al.'s study, an interval was perceived as longer when it was marked by two pulses delivered to different tactile locations (hands).

However, instead of being an effect of space, the results of Grondin et al. (2011) could be attributed to the latency to integrate two markers stimulating different cortical hemispheres. A stimulus to the left hand is projected to the right hemisphere and a stimulus to the right hand is projected to the left hemisphere. If there is some latency to integrate stimuli projected to different hemispheres relative to the same hemisphere, such latency might have resulted in longer perceived duration in Grondin et al.'s study.

This argument becomes plausible when one considers what was suggested by Gescheider (1966). Gescheider examined how temporal resolution is changed when an inter-onset interval is marked by two vibrotactile signals stimulating different fingers, instead of the same finger. A threshold was estimated in this study: when an interval between two markers was briefer than this threshold, participants could not perceive these markers as discrete events but rather perceived them as fused into a single event. Gescheider reported that the threshold remained the same (10 ms) regardless of whether two pulses were delivered to the same (index) or different (ring and index) fingertips of the same hand, while delivering two pulses to the index fingertips of different hands led to a higher threshold (12.5 ms). This might suggest that duration processing is not affected by the cortical locations of stimuli within the same hemisphere while it changes depending on whether two markers stimulate the same or different hemispheres.

The present experiment was conducted to determine whether inter-stimulus intervals would be perceived as longer when they were marked by two pulses stimulating different fingers (of the same hand) than the same finger. In each trial, two intervals to be compared were successively delivered to the left hand. Two pulses bounding the first interval were both delivered to the middle fingertip. The onset pulse of the second interval was delivered to the middle fingertip while the offset pulse was delivered to the middle or little fingertip. The angle between the middle and little fingers was approximately 20° or 40°. Note that in this procedure two pulses stimulated the same or different cortical regions but within the same hemisphere.

Method

Participants

Twelve participants, six males and six females aged 19-48 years, self-reporting being right-handers, were recruited. They were students or employees at Université Laval, except one who was an unemployed resident of Québec. They consented to their participation by signing a form approved by the institutional ethical committee, and received $20 CAN for their participation.

Apparatus and stimuli

Stimulus presentations were controlled by E-prime software installed in an IBM computer. Empty time intervals were marked by electrical pulses of 20 ms. To electrify each fingertip, two electrodes (a plus pole and a minus pole) were attached to both sides of the distal interphalangeal (Figure 1). These electrodes were covered with conductive gel and were fixed with medical tape. An experimenter calibrated the voltage before the beginning of each session; the voltage was fixed to a point at which participants reported perceiving stimuli clearly but without any discomfort.

Four pulses were successively presented in each trial, the first two marking the *standard* interval and the last two marking the *comparison* interval. These intervals were manipulated in terms of temporal duration from the offset of the preceding pulse to the onset of the following pulse (i.e., these intervals were inter-stimulus intervals). The standard interval was fixed at 500 ms while the comparison interval was varied from 300 to 700 ms in steps of 80 ms, resulting in six comparison intervals. The standard and the comparison were separated by an inter-stimulus interval of 2000 ms.

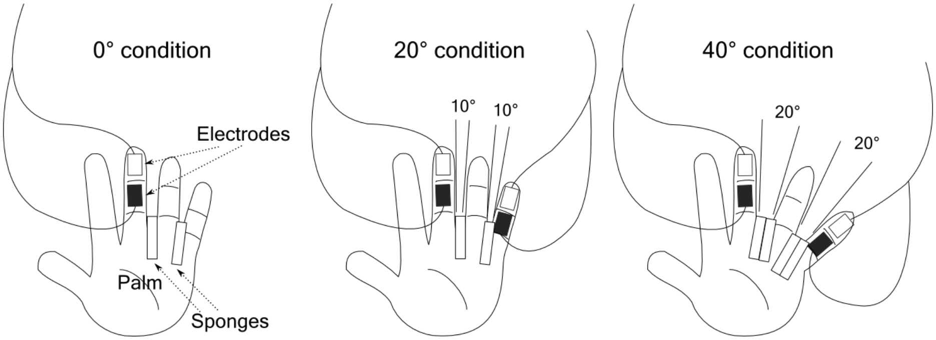


Figure 1. Electrodes' attachment.

All four pulses were delivered to the left hand. Two pulses bounding the standard were both delivered to the middle finger. The onset pulse of the comparison was delivered to the middle finger while the offset pulse of the comparison was delivered in one of the three conditions. In the 0° condition, the offset pulse was delivered to the middle finger, i.e., both onset and offset pulses were delivered to the same (middle) finger. In the 20° condition, it was delivered to the little finger, with the angle between the middle and little fingers being approximately 20°. In the 40° condition, it was delivered to the little finger, with the angle between the middle and little fingers being approximately 40°.

Angles between fingers were fixed with sponges (which were conventionally to be used as medical toe separators; see Figure 1). In the 0° and 20° conditions, one sponge was inserted between the middle and ring fingers and between the ring and little fingers. In the 40° condition, two sponges were inserted between the middle and ring fingers and between the ring and little fingers.

Procedure

Participants were instructed to judge whether the comparison interval was shorter or longer than the standard interval. Participants responded verbally and the response was registered by an experimenter who also stayed in the experimental booth. The next trial started 2 s after the registration of the response.

There were three sessions, each one being dedicated to one experimental condition. The order of these sessions was counterbalanced. Each session consisted of four blocks each containing 60 trials; the six comparison intervals were presented ten times each in random order. A break of about 30 s was taken between the blocks. Each session took about 30 minutes.

Results

The probability of perceiving a comparison as longer was calculated for each comparison interval, resulting in 6-point psychometric functions. Each point of functions was based on 40 responses (4 blocks × 10 repetitions). The cumulative normal distribution was fitted to the resulting curves. The goodness of fit was high: The R^2 value was above .90 in all 36 cases (12 participants × 3 experimental conditions).

There are two parameters of interest. The first one is the bisection point (BP), which is a point at which the function crossed 0.5. Note that this parameter here is used to express the perceived duration of the comparison interval. If the comparison interval is perceived as longer than the standard, participants respond “longer” more frequently, resulting in a downward shift of the BP. In other words, a lower BP indicates a longer perceived duration of the comparison interval. The second one is the standard deviation (SD) of the fitted cumulative normal distribution. Dividing one SD by 500 (the standard interval) gave the Weber fraction (WF), where a lower value indicates a steeper slope of the function and thus indicates better discrimination (Grondin, 2008). Individuals’ BP and WF are shown in Figure 2.

There were individual differences in the direction of spatial effect on their perceived duration (BP). Indeed, a repeated-measure analysis of variance (ANOVA) showed no significant difference for the pooled data, $F(2, 22) = 1.215$, $p = .316$, $\eta_p^2 = .100$. A cluster analysis based on squared Euclidian distance and Ward method divided participants into two groups (Figure 3). Note that before constructing the distance matrix for this analysis we standardized BPs *in each participant*, i.e., we calculated in each participant a mean BP across three experimental conditions, subtracted the mean from the BP of each condition, and divided the resulting value by a standard deviation across the three conditions. In eight participants, the comparison interval was perceived as longer, i.e., the BP was lower, when two pulses were delivered to different fingers than the same finger (the overestimation group). In four participants, the comparison interval was perceived as shorter when two pulses were delivered to different fingers (the underestimation group).

In each group, the 20° and 40° conditions led to near identical duration. The ANOVA showed that some difference was significant for the overestimation group, $F(2, 14) = 15.877$, $p < .001$, $\eta_p^2 = .694$, while the trend analysis revealed that the quadratic trend, $F_{quad}(1, 7) = 7.883$, $p = .026$, $\eta_p^2 = .530$, as well as the linear trend, $F_{lin}(1, 7) = 24.862$, $p = .002$, $\eta_p^2 = .780$, was significant. Indeed, multiple comparisons subjected to Ryan's adjustments showed that the 20° condition ($p < .001$) and the 40° condition ($p < .001$) led to longer perceived duration than the 0° condition while the difference between the 20° and 40° conditions was not significant ($p = .935$). For the underestimation group, there was no significant difference, $F(2, 6) = 3.204$, $p = .113$, $\eta_p^2 = .516$.

Discrimination (WF) was not affected by the location of pulses in both groups. The ANOVA showed no significant difference for the overestimation group, $F(2, 14) = .784$, $p = .476$, $\eta_p^2 = .101$, for the underestimation group, $F(2, 6) = 1.817$, $p = .242$, $\eta_p^2 = .377$, nor for the pooled data, $F(2, 22) = 2.359$, $p = .118$, $\eta_p^2 = .177$.

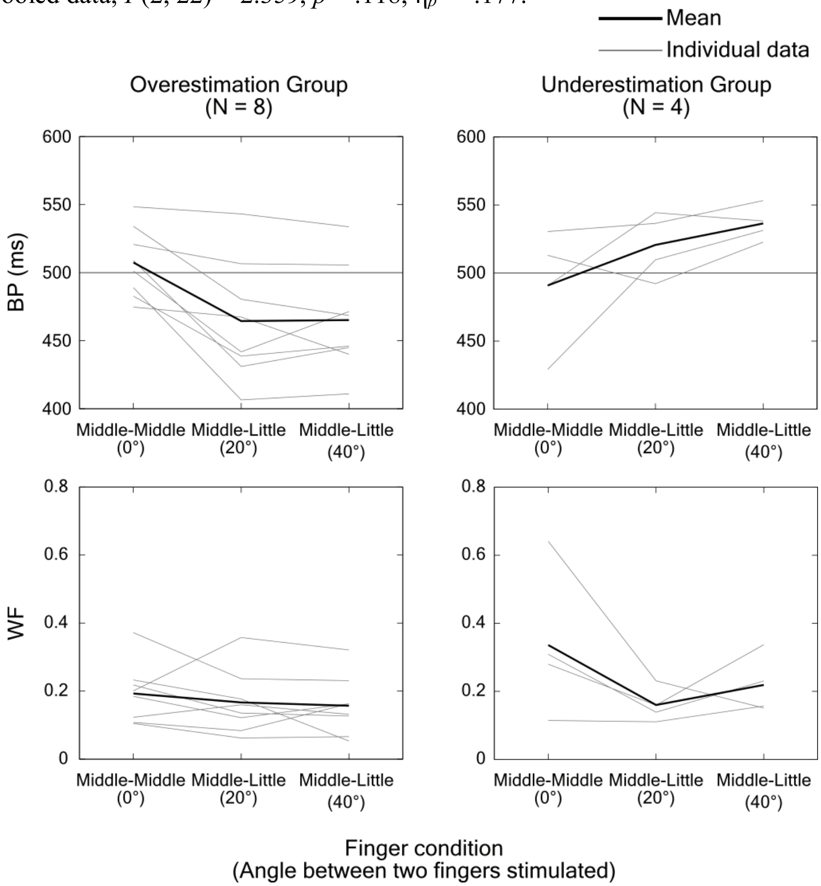


Figure 2. Bisection Point (BP: upper panel) and Weber Fraction (WF: lower panel) in each experimental condition for participants overestimating (left panel) and underestimating (right panel) the comparison intervals

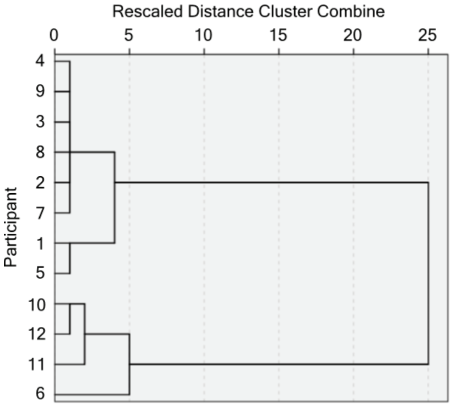


Figure 3. Dendrogram obtained by a cluster analysis on standardized BPs

Discussion

The results of the present study indicated that whether empty time intervals are overestimated or underestimated when they are marked by two pulses stimulating different fingers depends on individuals. Indeed, eight of twelve participants perceived empty time intervals as longer while four participants perceived intervals as shorter when two pulses were delivered to different fingers instead of to the same finger. Moreover, increasing the angle between the middle and little fingers did not magnify the spatial effect observed in both groups of participants; the 20° and 40° conditions led to near identical perceived duration in each group. This finding is consistent with that reported by Grondin et al. (2011). In their study, intervals were perceived as longer when two pulses were delivered to different hands than when presented to the same hand, while placing the two hands 3 feet apart resulted in almost the same duration as placing the hands nearby.

Grondin et al. (2011) delivered two pulses marking intervals to the same or different hands. In this procedure, two pulses stimulated the same or different cortical hemispheres. However, the present study revealed that there were individual differences in the perceived duration when two pulses stimulated different cortical regions within the same hemisphere. Given that intervals were overestimated even when the two hands were placed nearby in Grondin et al., delivering two pulses to different hemispheres seems a crucial factor causing the overestimation of duration.

The finding of the present study is partially inconsistent with that observed in our pilot experiment where participants perceived intervals as longer when the onset and offset markers were delivered to the index and ring fingers, respectively, compared to when both markers were delivered to the index finger. It is known that the index finger as well as the thumb has a better sensitivity for detecting stimuli than the other fingers (Johansson & Vallbo, 1979). Given that two pulses were delivered to the fingers of similar sensitivities in the present study, the overestimation observed in our pilot experiment is probably attributed to the fact that the index finger detects stimuli faster than the ring finger.

Temporal sensitivity (WF) remained almost the same regardless of whether participants overestimated or underestimated intervals marked by two pulses stimulating different fingers. Note that eight participants perceived intervals as significantly longer while their discrimination was not impaired when two pulses were delivered to different fingers. This is consistent with that observed in our pilot experiment, where discrimination remained the same even when the onset and offset markers were delivered to the index and ring fingers, respectively, compared to when both markers were delivered to the index finger. This could not be explained by utilizing the concept of an internal clock. It is widely accepted in the time perception literature that there is an internal clock for processing temporal information (see Grondin, 2001, 2010). This clock is usually assumed to be a pacemaker-counter device where the first module is reported to emit pulses that are accumulated by the second one. The amount of accumulation determines perceived duration; more pulses result in longer perceived duration. Moreover, more pulses result in more variability of accumulation, i.e., longer perceived duration leads to lower sensitivity. Indeed, in Grondin et al. (2011), delivering two pulses to different hands led to longer perceived duration and impaired sensitivity. However, in the present study, delivering two pulses to different fingers did not result in impaired sensitivity even when intervals were perceived as longer.

The absence of impaired sensitivity (by space between markers) in the present study is strikingly different from that reported by Grondin et al. (2011). In their study, discrimination was impaired when two pulses marking intervals were delivered to different hands relative to the same hand. However, Grondin et al.'s and the present studies each can be linked with the study of Gescheider (1966). In Gescheider's study, a) temporal resolution remained the same regardless of whether two pulses were delivered to the same or different fingers of the same hand, b) while resolution was impaired when two pulses were delivered to different hands. The former result (a) is consistent with that observed in the present study while the latter (b) is consistent with that reported by Grondin et al. In brief, temporal sensitivity is impaired when intervals are marked by two pulses stimulating different hands but remains the same when two pulses are delivered to different fingers within the same hand.

Acknowledgements

This research was made possible by a research grant awarded to SG by the Natural Sciences and Engineering Council of Canada.

References

Gescheider, G. A. (1966). Resolving of successive clicks by the ears and skin. *Journal of Experimental Psychology*, *71*, 378-381.

Grondin, S. (2001). From physical time to the first and second moments of psychological time. *Psychological Bulletin, 127,* 22-44.

Grondin, S. (2008). Methods for studying psychological time. In S. Grondin (Ed.). *Psychology of time* (pp. 51-74). Bingley, U.K.: Emerald Group Publishing.

Grondin, S. (2010). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception, & Psychophysics*, *72*, 561-582.

Grondin, S., Kuroda, T., & Mitsudo, T. (2011). Spatial effects on tactile duration categorization. *Canadian Journal of Experimental Psychology*, *65*, 163-167.

Johansson, R. S., & Vallbo, A. B. (1979). Detection of tactile stimuli. Thresholds of afferent units related to psychophysical thresholds in the human hand. *Journal of Physiology*, *297*, 405-422.

Suto, Y. (1952). The effect of space on time estimation (S-effect) in tactual space. I. *Japanese Journal of Psychology*, *22*, 189-200.

Suto, Y. (1955). The effect of space on time estimation (S-effect) in tactual space. II: The role of vision in the S effect upon the skin. *Japanese Journal of Psychology*, *26*, 94-99.

Yoblick, D. A., & Salvendy, G. (1970). Influence of frequency on the estimation of time for auditory, visual, and tactile modalities: The kappa effect. *Journal of Experimental Psychology*, *86*, 157-164.

INFLUENCES OF THE SIZES OF TACTILE BARS AND DOTS ON DISCRIMINABILITY IN PEOPLE WITH VISUAL IMPAIRMENT

Wataru Toyoda[1], Kouki Doi[2] and Hiroshi Fujimoto[3]
[1] Graduate School of Human Sciences, Waseda University
2-579-15 Mikajima, Tokorozawa, Saitama 359-1192, Japan
[2] Department of Teacher Training and Information, National Institute of Special Needs
Education, 5-1-1 Nobi, Yokosuka, Kanagawa 239-8585, Japan
[3] Faculty of Human Sciences, Waseda University,
2-579-15 Mikajima, Tokorozawa, Saitama 359-1192, Japan
wataru_toyoda@toki.waseda.jp, doi@nise.go.jp, fujimoto@waseda.jp

Abstract

Tactile bars and dots served as tactile landmarks so that people with visual impairment could use the same consumer products as those used by sighted people. However, reliable data on the appropriate sizes of products was not readiliy available. The purpose of this study was to evaluate the influence of tactile bar and dot sizes on the discriminability of the two among younger and older adults with visual impairment. This was done to determine appropriate tactile bar size (as distinguished from tactile dots). Participants tactually discriminated several sizes of tactile bars and dots presented individually, in random order, via a two-alternative forced-choice task (2AFC) using the index finger of the prefered hand. Results showed that longer dimensional differences between tactile bar width and length is an important factor for correctly discriminating tactile bars. Most participants in both groups distinguished tactile bars that were larger than +2.0 mm from tactile dots quickly and with high accuracy. Meanwhile, tactile dots with a larger edge radius of curvature had greater discriminability than tactile dots with a smaller edge radius of curvature in the case of dots with an identical diameter.

According to a survey conducted by the Japanese Ministry of Health, Labour and Welfare (2008) in 2006, there were about 300,000 people with visual impairment in Japan. It is important to market consumer products that meet the physical needs of individuals with visual impairment. Many countries across the world are also facing the same issue. Several consumer electronics manufacturers actively provide support to older adults by improving the tactile accessibility of their consumer products. In Japan, several companies apply tactile dots (dot-shaped tactile symbols) and bars (bar-shaped tactile symbols) to the manual operating portions of their existing products so that people with visual impairments can use the same products as those used by the visually able. The tactile bars and dots contribute greatly to the access of information and acceleration of independent communication of people with visual impairment. However, previously, many companies have used tactile bars and dots of various sizes more discretely; this has caused confusion for people with visual impairment. Under such circumstances, the Japanese Standards Association standardized new "Guidelines for all people including elderly and people with disabilities—Making tactile dots on consumer products" (JIS S 0011; 2000), to specify guidelines for the design of tactile bars and dots. Additionally, the International Organization for Standardization also enacted guidelines (ISO 24503; 2011) for tactile bars and dots on the basis of JIS S 0011.

The standards provide that tactile bars and dots be applied to manually operated product keys for two purposes: to provide location information for other functions on the device and to identify control functions. These standards also specify that tactile dots mark the button to start the basic function of the device and that tactile bars mark the button to stop the basic function. However, the recommended sizes of bars and dots were not determined by objective and quantitative data, and sufficient reliable data on their appropriate sizes is not readily available. Specifically, the standards do not contain provisions concerning cross-sectional shapes, despite the probablility that cross-sectional shapes affect the operational performance of these products. Hence, quantitative evidence for the appropriate dimensions of tactile bars and dots is required to revise existing (and devise new) standards. One of the most important issues is to clarify the sizes of tactile bars as distinguished from tactile dots among users with visual impairment. This is because tactile bars and dots are applied to keys with different functions as mentioned above.

In this paper, we evaluated the influence of tactile bar and dot sizes on the discriminability of the two among people with visual impairment.

Method

Participants

Ten younger adults (mean age 26.1±4.6 years; range of duration of vision loss 0–3 years old) and 10 older adults (mean age 65.3±4.0 years; range of loss of vision 0–12 years old) with visual impairment participated in this experiment. All participants were without skin injuries or dermatosis on their upper limbs.

Stimuli

Dots and bars of several sizes were used as stimuli for this experiment. Dot diameter, bar width, bar length, and the edge radii of their curvature (hereafter referred to as the "R") were controlled (refer to Fig. 1 and Table 1). The height of all stimuli was 0.5 mm.

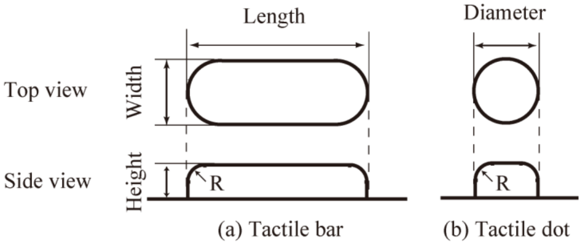


Fig. 1. Dimensions of tactile bars and dots used in the experiment

Table 1. Size conditions of tactile bars and dots

(a) Tactile bar conditions

R (mm)	Length (mm) *	Width (mm)
0.00, 0.25	+0.5, +1.0, +2.0, +3.0, +4.0	0.5, 0.8, 1.0, 1.5, 2.0
0.50	+0.5, +1.0, +2.0, +3.0, +4.0	1.0, 1.5, 2.0

(b) Tactile dot conditions

R (mm)	Diameter (mm)
0.00, 0.25	0.5, 0.8, 1.0, 1.5, 2.0
0.50	1.0, 1.5, 2.0

* The actual length is the size that adds each width to each length condition.

Tactile bar conditions were as follows. Bar widths were 0.5 mm, 0.8 mm, 1.0 mm, 1.5 mm, and 2.0 mm with each R of 0.0 mm and 0.25 mm. Moreover, the widths were 1.0 mm, 1.5 mm, and 2.0 mm with an R of 0.5 mm. We controlled the length of all tactile bars, and we added each width to each length condition: +0.5 mm, +1.0 mm, +2.0 mm, +3.0 mm, +4.0 mm. There were 65 conditions for the bar stimuli.

Tactile dot conditions were as follows: the diameters were 0.5 mm, 0.8 mm, 1.0 mm, 1.5 mm, and 2.0 mm with each R of 0.0 mm and 0.25 mm. The diameters were 1.0 mm, 1.5 mm, and 2.0 mm with an R of 0.5 mm. There were 15 conditions for the dot stimuli.

All the above-mentioned conditions covered current recommended standardized sizes of tactile bars and dots. All stimuli were created by cutting the acrylic at the center of a 50 mm × 50 mm acryl plate.

Procedure

Participants tactually discriminated stimuli presented individually, in random order, by a two-alternative forced-choice task (2AFC, bar or dot). Participants touched the stimuli using only their index fingertip of the preferred hand throughout this experiment. Sixty-five bar conditions were presented three times each, and 13 dot conditions were presented 15 times each for a total of 390 trials for each participant (so as to eliminate order effects). Practice trials were presented before starting the experiment. Accuracy rates and discrimination times were measured. The influences of diameter and the R of tactile dots within each participant group were evaluated using analysis of variance (ANOVA) followed by Bonferroni adjustments for multiple comparisons on accuracy rates and discrimination times.

This study obtained permission from the Ethical Committee on Human Research of Waseda University.

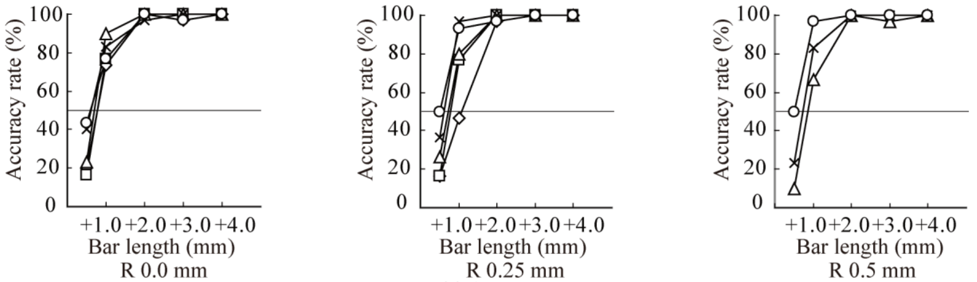
Results and Discussion

Results of the tactile bar conditions

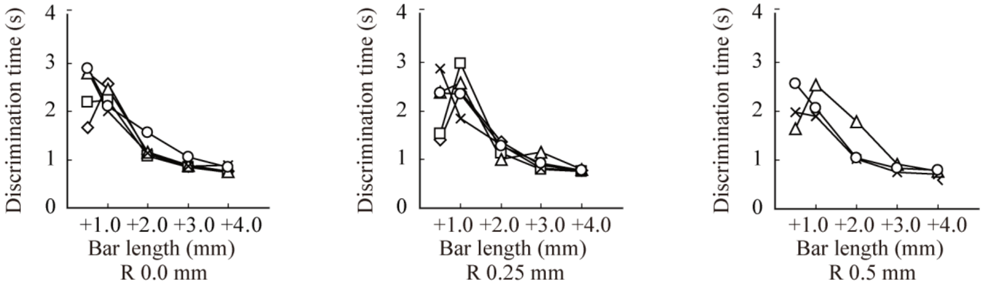
Both participant groups discriminated tactile bars from tactile dots faster and more accurately as the dimensional difference between bar length and width increased, regardless of the width and R conditions (refer to Fig. 2 & 3). In particular, most participants in both groups distinguished tactile bars of more than +2.0 mm from the tactile dots in less time and with high accuracy. On the other hand, in the + 0.5 mm and + 1.0 mm length conditions, both groups were not able to correctly discriminate all sized conditions of the tactile bars.

Previous studies have shown that both sighted individuals and those with visual impairment display age-related declines in tactile spatial resolution when passive tactile stimuli are applied to their index finger (Stevens et al., 1996; Goldreich et al., 2003). In contrast, Legge at el. (2008) proposed that people with visual impairment retain high spatial acuity of active touch well into old age on account of their daily tactile activities. Therefore, there was no marked difference in the tactile bar length condition between the two age groups because older participants had rich experiences with using touch for approximately 50 years since the loss of their vision.

On instrumental grounds, it is important to note that among younger and older adults with visual impairment, the discriminability of tactile bars depends on the difference between their length and width. The width and the R of the bars were not critical factors. Consumer electronics manufacturers cannot necessarily control the edge radius of curvature of tactile bars owing to limitations in the manufacturing process. However, they can apply

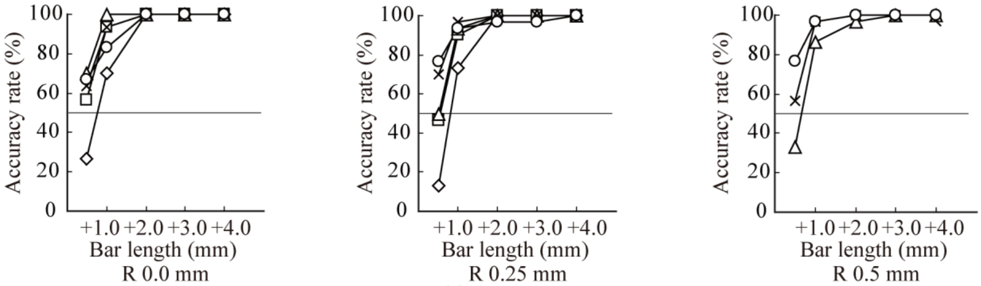


(a) Accuracy rate

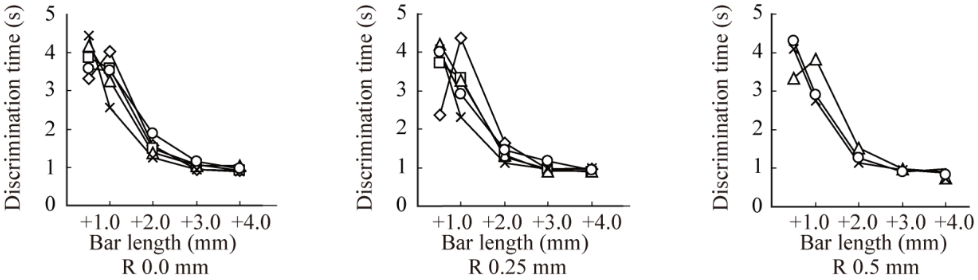


(b) Discrimination time

Fig. 2. Results of tactile bar conditions among younger participants



(a) Accuracy rate



(b) Discrimination time

Fig. 3. Results of tactile bar conditions among older participants

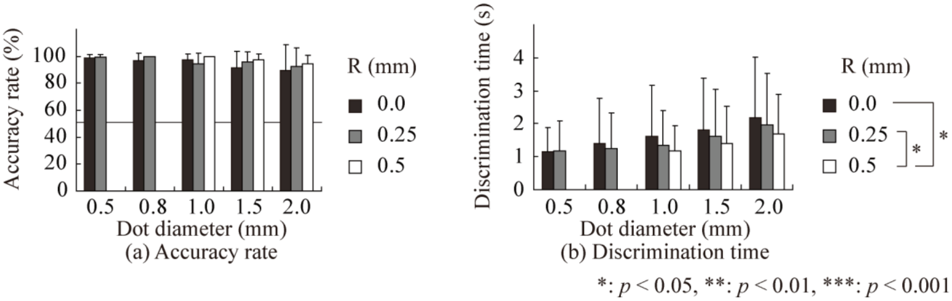


Fig. 4. Results of tactile dot conditions among younger participants

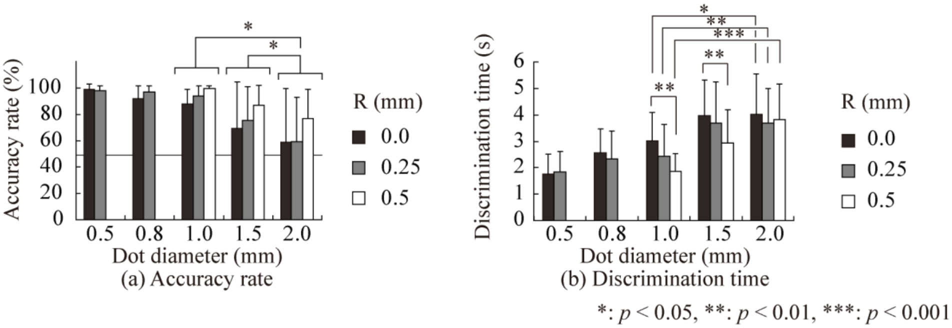


Fig. 5. Results of tactile dot conditions among older participants

discriminable tactile bars to their products by designing the bars in a way that they contain a sufficient difference between their width and length.

Results of the tactile dot conditions

The ANOVA results revealed the following effects. There was a significant effect of discrimination times in the younger group for diameter ($F_{2,\ 18} = 11.11$, $p < .001$). There was also a main effect for the R ($F_{2,\ 18} = 6.78$, $p < 0.05$) and diameter ($F_{2,\ 18} = 10.03$, $p < .01$) in accuracy rates for older adults. Each R ($F_{2,\ 18} = 13.86$, $p < .001$) and diameter ($F_{2,\ 18} = 20.50$, $p < .001$) as well as their interaction ($F_{4,\ 36}) = 3.42$, $p < .05$) in terms of discrimination times achieved statistical significance. Overall, tactile dots with a larger R had higher discriminability than those with a smaller R for dots that were of identical height. Furthermore, both groups tended to perceive a tactile bar with a large R as a tactile dot when the difference between the length and width of the bar was very small, as in the +0.5 mm length condition (refer to Fig. 4 & 5).

One possible reason for the high discriminability of tactile dots with a large R is the loss of the dots' morphological characteristics. Participants might perceive tactile dots with large a R as mere point stimulations, because the dots do not have sharp edges that are easily traceable by the finger nor do they have a large, flat surface on the top. Additionally, the superficial, touchable dimensions of these tactile dots might be slightly smaller than those dots with a small R. This is because the former dots have wide skirts, and skin tissue cannot easily conform to them. Thus, participants might perceive the dots with a large R as smaller than they really are. For these reasons, tactile dots with a larger R are more suitable when a tactile bar and a tactile dot are concurrently applied.

Acknowledgements

This work was supported by a Grant-in-Aid for Scientific Research (B) (Number 22300202) and Global COE Program “Global RobotAcademia” from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

Ministry of Health, Labour and Welfare in Japan. (2008). Survey on persons with pyhsical disability in 2006, Ministry of Health, Labour and Welfare in Japan (Eds.), Ministry of Health, Labour and Welfare in Japan.

Japanese Standards Association. (2000). JIS S 0011: Guidelines for all people including elderly and people with disabilities - Making tactile dots on consumer products, Japanese Standards Association, Tokyo, (In Japanese)

International Organization for Standardization. (2011). ISO/DIS 24503: Ergonomics - Accessible design - Tactile dots and bars on consumer products, International Organization for Standardization, Geneva.

Goldreich D., Kanics I.M. (2003). Tactile acuity is enhanced in blindness, *Journal of Neuroscience*, 23, 3439-3445.

Stevens J.C., Foulke E., Patterson M.Q. (1996). Tactile acuity, aging, and Braille reading in long-term blindness, *Journal of Experimental Psychology*, 2, 91-106.

Legge G.E., Madison C., Vaughn B.N., Cheong A.M., Miller J.C. (2008). Retention of high tactile acuity throughout the life span in blindness, *Perception & Psychophysics*, 70(8), 1471-1488.

THE PRESENTATION OF LONG TERM DURATION OF BODY MOVEMENT IN IMPRESSIONIST ARTWORKS DIFFERENTLY DISTORT THE PERCEPTION OF TIME

Francisco Carlos Nather and José Lino Oliveira Bueno
Department of Psychology, University of São Paulo, Faculty of Philosophy Sciences and Letters of Ribeirão Preto, São Paulo, Brazil
f.nather@hotmail.com / jldobuen@ffclrp.usp.br

Abstract

This work verified whether images implying movement exposed for fixed different durations affect the perception of time. Undergraduate participants observed pictures of sculptures of dancers by Edgar Degas for 9, 18, 27 or 45 seconds (G9, G18, G27 and G45 groups, respectively) and the stimuli were randomly presented in arithmetical (1.5-, 3.0- and 4.5-point) or geometrical (1.5-, 3.0- and 6.0-point stimuli) progressions. The reproduction method to record the time estimations of the subjects was used. Data analysis showed that time was not distorted in the G9, G18 and G45 groups, except: 6.0-point stimulus was overestimated in geometrical (G9) and 1.5-point was underestimated in arithmetical (G45) progressions. However, time distortions in the G27 group were modulated by different implied movement intensities as was observed in previous works that used 36 s of image exposure. These results show that different processes involving the visual perception of movement in static images are also associated to the different exposure duration.

Time perception is surprisingly subject to distortions. Temporal judgments are constructions of the brain which are also easy to experimentally manipulate (Eagleman, 2008). Cognitive functions such as attention, working memory and long term memory determine temporal judgments (Brown 1997; Wittmann, 2009).

Time distortions have been related to methodological procedures, different paradigms and the *duration* of stimuli. Shorter intervals would be more sensory in nature (Block, 1989; Beher, Desjardins, & Fortin, 2007). Durations of milliseconds to few seconds (2-3 s) have been related to the biological models of the subjective time. Intervals close to 3 s tend to be estimated accurately, whereas longer intervals (above 5 s) are not only related to the sensorial perception and tend to be distorted (Ulbrich, Churan, Fink, & Wittmann, 2007; Noulhiane, Pouthas, & Samson, 2008).

Longer intervals seem to be more cognitively mediated (Block, 1989; Zakay & Block, 1997). Therefore, durations of several seconds and minutes were used in researches focusing not only on the sensorial perception but also on the cognitive processes of time perception and, because of this, related to cognitive models of the subjective time.

Static images of short (milliseconds) and long (above 5 s) durations have been used to explain different aspects of time perception. For example, pictures of objects such as shoes, houses and fruits were used in studies with short duration (milliseconds). They showed that an unexpected image affect time perception (Eagleman & Pariyadath 2009). Pictures of people exposed for 2, 4 and 6 s caused time distortions related to the arousal levels evoked by "positive" and "negative" emotional contents of the images (Angrili, Cherubini, Pavese, & Manfredini, 1997).

Artworks were used to test the effects of pleasure on time perception (Cupchick & Gebotys, 1988). Using impressionist paintings the authors showed that time distortions reflected the interpretive activity evoked by artworks when the subjects observed the paintings for 36 s. Furthermore, time distortions were also related to the duration of exposures: when the paintings were exposed for 18 s and 72 s they were respectively over- and underestimated. However, the authors did not inform what aspects (characteristics) of the paintings were related to these time distortions.

Long duration exposure of 36 s was used in several studies that verified the effect of body movements in static images on subjective time perception (e. g., Nather & Bueno, 2006, 2011). In these researches two, three or four pictures of Edgar Degas' ballerinas in different ballet steps were randomly presented to the subjects. Using the reproduction method, these researches revealed that the processing of motion experience modulates the perception of time generating under- and overestimations. Further study using the same ballerina images in the bisection method showed that short exposures of 0.4 to 1.6 s strongly affected the perception of time than exposures of 2 to 8 s (Nather, Bueno, Bigand, & Droit-Volet, 2011). These works confirmed that an important characteristic of a static image – implied movement – affects time perception independently of the duration of exposure.

Subjective time literature had pointed that subjects tend to overestimate the short intervals and underestimate the long ones. Lewis and Miall (2009) used different time exposure to verify how the duration of stimuli affect time perception in terms of scalar property of interval timing. They studied time across a broad range of intervals (milliseconds to minutes) to support the hypothesis that distinct clock mechanisms are used for different subsets of time rankings. They demonstrated a gradual increase in precision of timing as the intervals measured increase (see also Bizo, Chu, Sanabria, & Killen, 2006; Merchant, Zarco, & Prado, 2008).

Several subjective time experiments have used static images of different semantic contents and durations to verify, for example, the effects of pleasure and emotions on time perception (e. g., Angrili et al., 1997; Droit-Volet & Meck, 2007; Droit-Volet & Gil, 2009). However, the use of static images having implicit scalar property (implied movement) presented by different durations would inform important aspects of subjective time perception. The aim of the present study was to verify whether the induced movements by different body postures affect time perception when they are presented for different longer durations (9, 18, 27 and 45 s). Therefore, Degas' ballerinas were used for ranking 1.5 to 6.0 points of induced movement according to the Body Movement Ranking Scale - BMRS (Nather & Bueno, 2008).

Method

The experiment was approved by the Ethics Committee of the University of São Paulo School of Philosophy, Sciences, and Letters in Ribeirão Preto, Brazil.

One hundred and twenty-two undergraduate students (50 men; M age = 20.94 yr., SD = 2.58) untrained in visual arts and ballet dance from University of São Paulo of Ribeirão Preto participated of the experiment. They reported having normal or corrected-to-normal vision. The experiment was performed during daylight hours in an isolated, soundproofed room at the University of São Paulo. The lights were off during the experiment.

Digital photographs of four ballerinas sculptures by Edgar Degas were used as stimuli: "Ballerina in Repose with her Hands on the Waist and her Left Leg in Front" (facing forward) (1.5-point), "Ballerina in Repose, with her Hands on the Waist and her Right Leg in Front" (facing to the right) (3.0-point), and "Spanish Dance" (4.5-point), and "Third Stage of the Great Arabesque" (6.0-point) (Figure 1).

Exposure of stimuli and recording of time estimations were done by *Wave Surfer* program installed on an HP notebook. The tasks were orally explained to the participants. They were positioned facing the central region of the LG 19” monitor and were asked not to count time. The stimuli were exposed by pressing the “presentation” key and their time exposure were finalized after 36 s. At this moment, the monitor was filled with white color indicating that the participant could initiate time estimation. Then, immediately after each time observation the participant reproduced the presentation duration by pressing the “initiate” key. The experienced duration of each stimulus was finalized by pressing the “finished” key (reproduction method).

The stimuli were presented to four groups of participants according to its fixed time of exposure of 9, 18, 27 or 45 s: G9 (n=30), G18 (n=30), G27 (n=31) and G45 (n=31) groups, respectively. Three stimuli grouped in arithmetical (1.5-, 3.0- and 4.5-point stimulus) or geometrical sequences (1.5-, 3.0- and 6.0-point stimulus) were randomly presented to each group in a manner that half of the participants of each group observed only one of these two progressions orders. The sequences of stimuli presentation were randomly ordered. Training stimulus (4.0-point) was presented before each sequence in order to make the participant familiar with the experimental task but these data were not used in the analyses.

1.5-point **3.0-point** **4.5-point** **6.0-point**

Figure 1. The four sculptures (1.5-, 3.0-, 4.5- and 6.0-point stimuli) used in the experiment. The stimuli were obtained from the “Body Movement Ranking Scale” (Nather & Bueno, 2008). © Edgar Degas. Paris, France 1834-1917, MASP Collection, Museum of Art of São Paulo Assis Chateaubriand. Pictures of João L. Musa.

One Way (ANOVA) analyses without repeated measures were used to compare time estimation data in the G9, G18, G27 and G45 groups in both arithmetical and geometrical progressions. One Way (ANOVA) analyses were also used considering the mean values of the two progressions (see italics, Table 1). Student-Newman-Keuls test was used for *post hoc* comparison of these analyses. Student *t* test was used to compare the time estimations data with the actual time of exposures of 9, 18, 27 and 45 s of G9, G18, G27 and G45 groups respectively.

Results

Mean values of time estimations of the G9, G18, G27 and G45 groups are presented in Table 1. The time estimations analysis (ANOVA) of Groups G9 and G45 did not show a significant main effect between stimuli in both arithmetical and geometrical progressions. The analysis of mean (M) progressions of these groups presented similar result. However, in relation to the real time of exposition the 1.5-point stimulus in the G45 group was underestimated in arithmetical progression [*t*(30)=2.29; *p*<.05] and considering the mean values of the two progressions [*t*(60)=2.98; *p*<.01]; the 6.0-point stimulus in the G9 group was overestimated in geometrical progression [*t*(28)=-2.01; *p*<.01]. Group G18 analyses did not show any temporal distortions.

Data analyses of Group G27 revealed distinct time distortions among stimuli in the geometrical progression [$F(2, 42)=6.91$; $p<.01$] and considering the mean values of the two progressions [$F(3, 89)=4.63$; $p<.01$]. *Post hoc* comparisons showed that in geometrical progression the 1.5- and 3.0-point stimuli were estimated shorter than 6.0-point stimuli; and considering the mean values of the two progressions the 1.5- and 3.0-point stimuli were estimated shorter than both 4.5- and 6.0-point. In both analyses 1.5- and 3.0-point were not statistically different, as well as 4.5- and 6.0-point were not different considering the mean values of the two progressions. These results showed that the stimuli inducing less movement were estimated with shorter duration than those inducing greater movements.

The comparisons of stimuli time estimations with the actual time of 27 s of image exposures showed that 1.5- and 3.0-point stimuli were underestimated in geometrical [$t(28)=2.02$; $p=.05$] and [$t(28)=3.05$; $p<.01$] and considering the mean values of the two progressions [$t(60)=2.471$; $p<.01$] and [$t(60)=1.98$; $p<.05$]. On the contrary, the 6.0-point stimuli were overestimated [$t(28)=-2.26$; $p<.05$].

Table 1. Mean values and standard errors (brackets) of the participants' temporal estimations for 1.5-, 3.0-, 4.5- and 6.0-point stimuli from the BMRS (Nather & Bueno, 2008) of Groups G9 (n=30), G18 (n=30), G27 (n=31) and G45 (n=31). Values are expressed in seconds.

Group	Progression	1.5-point M (SD)	3.0-point M (SD)	4.5-point M (SD)	6.0-point M (SD)
G9	Arithmetical	8.93 (1.80)	9.10 (2.85)	9.58 (1.94)	
	Geometrical	9.43 (1.81)	10.28 (2.45)		11.05 (2.54)+
	Mean (M)	*9.18 (1.79)*	*9.69 (2.68)*	*9.58 (1.94)*	*11.05 (2.54)+*
G18	Arithmetical	18.62 (5.84)	17.61 (5.36)	18.11 (5.92)	
	Geometrical	18.40 (2.69)	17.70 (4.06)		18.73 (3.51)
	Mean (M)	*18.51 (4.47)*	*17.66 (4.66)*	*18.11 (5.92)*	*18.73 (3.51)*
G27	Arithmetical	25.93 (2.70)	26.69 (4.67)	28.97 (6.27)	
	Geometrical*	24.40 (4.96) -	24.14 (3.62) -		29.11 (3.61)+
	*Mean (M)***	*25.19 (4.06) -*	*25.46 (4.33) -*	*28.97 (6.27)*	*29.11 (3,61)+*
C45	Arithmetical	40.75 (7.40) -	44.26 (12.21)	44.26 (10.28)	
	Geometrical	42.20 (5.84)	42.20 (5.92)		43.79 (8.10)
	Mean (M)	*41.45 (6.62) -*	*43.26 (9.59)*	*44.26 (10.28)*	*43.79 (8.10)*

Note: The values of the 4.5- and 6.0-point stimuli were repeated because these stimuli were presented once in the arithmetical or in the geometrical progression. (-) Underestimated; (+) Overestimated; (*) Indicate 1.5 = 3.0 < 6.0; and (**) Indicate 1.5 = 3.0 < 4.5 = 6.0.

Discussion

Presented for different durations, the images of body positions inducing distinct movements altered the subjective time perception. The results of 27 s of exposure agreed with those of previous studies using fixed duration of 36 s of exposure (Nather & Bueno, 2006, 2011). These researches emphasized that subjective time was modulated by induced movement: 1.5- and 3.0-point (less movement) were estimated shorter than 4.5-point (intermediate movement) and 6.0-point (higher movement). Also, 1.5- and 3.0-point stimuli were underestimated and

6.0-point stimulus was overestimated, but 4.5-point stimulus was accurately estimated (see Table 1).

On the other hand, modulatory effect of induced movement on time perception was not confirmed for 9, 18 and 45 s exposures. However, an interaction effect between time exposures and progressions of stimuli was observed: 6.0-point stimulus was overestimated in geometrical progression for 9 s of duration and 1.5-point was underestimated in arithmetical progression in 45 s duration.

The effects of duration in artworks were previously related (Cupchick & Gebotys, 1988). Using the same impressionist paintings, the authors verified overestimations when they were presented for 18 s. On the other hand, they pointed out that exposures for 36 s allowed the subjects to process the pictorial characteristics of the paintings in which generated time estimations related to the content of paintings. Because of this, they explain that longer exposures (72 s) caused underestimations due to the decrease of interest by the subjects.

This study was conduced by using the reproduction method in the prospective paradigm of subjective time as in previous studies using 36 s durations (Nather & Bueno, 2006, 2011). Taking this into account, it is possible to infer that the induced movement was more interactive with the durations ranging at 27 and 36 s. Why did the time distortions, related to the implied movement, occurred effectively at these durations of exposures?

According to Eagleman (2008), short and long durations can explain different biological and cognitive processes involved in internal time, because time distortions can be induced by properties of the stimuli themselves. Several cognitive processes, such as working memory, long-term memory, attention and decisions are involved in prospective time perception and different neural systems are involved in temporal processes depending on the duration of the processed duration (Wittmann, 2009). For example, Nather and Bueno (2012) used a methodological procedure (exploration time) in which the subjects were allowed to observe the images for any length of time and, immediately after each image was observed, recorded the duration as they perceived it. They verified that in mean, the subjects observed the images of Degas' ballerinas for 18 s and overestimated the images representing more movement (4.5- and 6.0-point stimuli).

This study used longer durations. It is possible to point out that long intervals evoke adjustments in subjects between the physical passing time and the time experienced internally according to the interaction between the procedure adopted and stimuli durations. Thus, the synchronization of real time and experienced time would involve different strategies of attention and sources of memory.

Artworks exposed for different durations in different methodological procedures have confirmed that body movement representation in static artwork images affect time perception. However, the time distortions were not totally coincident because of the time durations of exposures and procedures of the experiments (see Nather & Bueno, 2011, 2012; Nather et al., 2011). From this perspective, it was pointed out the necessity to run more parametric experimental studies involving systematic comparisons of time perception across different task conditions and different durations.

Acknowledgements

Francisco C. Nather received a Postdoctoral Scholarship and a Young Investigators Awards support both from São Paulo Research Foundation (FAPESP). José L. O. Bueno received a Research Grant from the National Counsel for Technological and Scientific Development.

References

Angrilli A., Cherubini P., Pavese A., & Manfredini S. (1997). The influence of affective factors on time perception. *Perception & Psychophysics*, *59*, 972-982.

Block, R. A. (1989). Experiencing and remembering time: Affordances, context, and cognition. In I. Levin & D. Zakay (Eds.), *Time and human cognition: A life span perspective* (pp. 333–363). Amsterdam: Holland.

Bizo, L. A., Chu, J. Y., Sanabria, F., & Killeen, P. R. (2006). The failure of Weber's law in time perception and production. *Behavioral Processes*, *71*, 201–210.

Bherer, L., Desjardins, S. & Fortin, C. (2007). Age-related differences in timing with breaks. *Psychology and Aging*, *22*, 398–403.

Brown, S. W. (1997). Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. *Perception and Psychophysics*, *59*, 1118-1140.

Cupchick, G. C., & Gebotys, R. (1988). The experience of time, pleasure, and interest during aesthetic episodes. *Empirical Studies of the Arts*, *6*, 1-12.

Droit-Volet, S., & Meck, W. H. (2007). How emotions colour our time perception. *Trends in Cognitive Sciences*, *11*, 504-513.

Droit-Volet, S., & Gil, S. (2009) The time-emotion paradox. *Journal of Philosophical Transactions of the Royal Society*, *B364*, 1943-1953.

Eagleman, D. M. (2008). Human time perception and its illusions. *Current opinion in Neurobiology*, *18*, 131-136.

Eagleman, D. M., & Pariyadath, V. (2009). Is subjective duration a signature of coding efficiency? *Journal of Philosophical Transactions of the Royal Society*, B364, 1841-1851.

Lewis, P.A., & Miall, R.C. (2009) The precision of temporal judgment: milliseconds, many minutes, and beyond. *Philosophical Transitions of the Royal Society*, *B364*, 1897-1905.

Merchant, H., Zarco, W., & Prado, L. (2008). Do we have a common mechanism for measuring time in the hundreds of millisecond range? Evidence from multiple-interval timing tasks. *Journal of Neurophysiology*, *99*, 939–949.

Nather, F. C., & Bueno, J. L. O. (2006). Efeito de imagens estáticas com diferentes representações de movimento sobre a percepção subjetiva de tempo. *Psicologia: Reflexão & Crítica*, *19*, 217-224.

Nather, F. C., & Bueno, J. L. O. (2008). Movement ranking scale of human body static images for subjective timing estimation. In *Proceedings of the Annual Meeting of the International Society for Psychophysics*, *24*, 185-190.

Nather, F. C., & Bueno, J. L. O. (2011). Static images with different induced intensities of human body movement affect subjective time. *Perceptual and Motor Skills*, *113*, 157-170.

Nather, F. C., & Bueno, J. L. O. (2012). Exploration time of static images implying different body movements causes time distortions. *Perceptual and Motor Skills*, *115*, 105-110.

Nather, F. C., Bueno, J. L. O., Bigand, E., & Droit-Volet, S. (2011). Time changes with the embodiment of another's body posture. *PLoS ONE*, *6*, e19818.

Noulhiane, M., Pouthas, V. & Samson, S. (2008). Is time reproduction sensitive to sensory modalities? *European Journal of Cognitive Psychology*, *21*, 18–34.

Ulbrich, P., Churan, J., Fink, M., & Wittmann, M. (2007). Temporal reproduction: Further evidence for two processes. *Acta Psychologica*, *125*, 51–65.

Wittmann, M. (2009). The Inner experience of time. *Journal of Philosophical Transactions of the Royal Society*, *B364*, 1955-1967.

Zakay, D., & Block, R. A. (1997). Temporal cognition. *Current Directions in Psychological Sciences*, *6*, 12-16.

TIMING PERCEPTION IS AFFECTED BY CUBIST PAINTINGS REPRESENTING HUMAN FIGURES

Francisco Carlos Nather, Paola Alarcon Monteiro Fernandes and José Lino Oliveira Bueno
Department of Psychology, University of São Paulo, Faculty of Philosophy Sciences and Letters of Ribeirão Preto, São Paulo, Brazil
f.nather@hotmail.com / jldobuen@ffclrp.usp.br

Abstract

The perception of time can be affected by different pictorial characteristics of artworks. Overestimations, related to higher levels of arousal, were obtained when people observed figurative artworks representing human body movements exposed for approximately 1.5 s. As the levels of arousal are maintained for 2-3 s, this study verified whether abstract paintings representing different types of movement affect time perception. Undergraduate participants observed 20 paintings implying different types of movement that were exposed for 3 s in random sequences. The prospective paradigm to record the time estimations was used. Data analysis showed that cubist paintings representing human figures were differently perceived: the painting with greater arousal and more implied movement was estimated longer than the paintings with less arousal and movement. These results are in agreement with those that used figurative human bodies exposed for short and long term durations which were explained through embodiment mechanisms.

Movement representation in abstract and figurative artworks has been discussed in the scientific literature (Braddick, 1995; Cutting, 2002). According to Cutting (2002), the painters use at least five different criteria to represent motion, such as to blur the background of the scene (photographic blur) and superimposing different instants of a moving object in the same painting (stroboscopic images). For example, in the abstract cubist painting *Nude Descending a Staircase* by Marcel Duchamp the human figure is repeatedly overlapped in order to show all the movements of the human body descending the stairs. Similarly, the futurist painting *Dynamism of a Dog on a Leash* by Giacomo Balla show to the viewers the movements done by the dog using stroboscopic images and also by blurring the background of the scene.

Kim and Blake (2007), using 24 paintings of the early 20th century, showed that the subjects were able to distinguish the movement intensity of different artworks: the paintings portraying motion were rated with higher scores than the paintings in which motion is not explicitly intended by artists. They found greater activity for the MT+ encephalic brain area in *Nude Descending a Staircase* and in *Dynamism of a Dog on a Leash* paintings (more movement) than in abstract expressionist paintings (less movement) in observers with prior viewing experience of these artworks. In these experiments, the authors used expressionist paintings made by Paul Klee and Piet Mondrian that are basically composed of geometrical forms and lines.

On the other hand, Thakral, Moo and Slotnick (2012) evaluated the magnitude of activity in sensory motion processing region MT+ and also in the prefrontal cortex, while subjects viewed 20 post impressionist paintings: the portraits and the landscapes of van Gogh. Based on each subject's motion experience ranking, the authors showed a significant positive correlation between the intensity of movement associated to viewing each painting and the magnitude of activity in the MT+ area. Furthermore, the paintings classified as pleasant were associated with greater activity in MT+, in relation to the paintings classified as unpleasant.

These results suggested that the activity of this brain area is associated with the experience of both movement and aesthetical pleasure.

The effects of aesthetical episodes on time perception has revealed that the experience of duration varies directly with the processing of the pictorial characteristics of artworks. They determine how long an observer spends on cognitive activities to understand a piece of art (Cupchick & Gebotys, 1988). Using paintings by Claude Monet, Edgar Degas and Édouard Manet exposed for 18, 36 and 72 s, these authors found that judgments of pleasure by subjects varied according to the complexity and the exposure duration of these artworks. However, this work was not focused on the effect of movement perception on subjective time.

Photographs of ballerina sculptures by Edgar Degas in different ballet postures, representing movements of distinct intensities, modulate time perception (Nather & Bueno, 2011). Presented for 36 s, the static postures (ballet steps) with lower movement scores were judged to have shorter duration than those with intermediary scores, and the ballet postures with intermediary scores were judged to be shorter than those with the highest scores (*arabesque* and *attitude*). According to the authors, arabesque and attitude postures are more asymmetric than other static ballet postures - because of the relationships among the body parts positions - providing to the observer subsides to understand the represented movements (see Cutting, 2002). Probably, the realism of the body postures of Degas's sculptures affect MT+, as was observed in abstract and figurative paintings (Kim & Black, 2007; Thakral, Moo, & Slotnick, 2012), since this brain area was differently activated by real pictures of people performing different movements: the observation of an athlete throwing a basketball cause more activation of MT+ than this same athlete in a static posture (Kourtzi & Kanwisher, 2000). Moreover, an element of aesthetic episodes consists of the activation of specific mechanisms related to the simulation of actions and corporeal sensation that is being observed by embodiment (Freedberg & Gallese, 2007).

Embodiment mechanisms have been associated with time distortions. Nather, Bueno, Bigand, and Droit-Volet (2011) showed that the duration was judged longer for Degas' ballerinas requiring more movement than for the ballerinas requiring less movement. However, the magnitude of the time distortions was relatively greater for the shorter (0.4-1.6 s) than for the longer (2-8 s) durations of stimuli exposures. Further, the authors found a positive relation between the intensity of movement and arousal levels: body postures requiring more movement were judged with more arousal than body postures requiring less movement.

Using pictures of people of different emotional contents which were presented for 2, 4 and 6 s, Angrili, Cherubini, Pavese and Manfredini (1997) verified that arousal levels were associated with time distortions. They showed that the subjects underestimated all the pictures as the arousing effect is transient and limited to brief durations shorter than 2-3 s.

Time perception was affected by figurative artworks exposed for different durations. Different characteristics (movement, pleasure and arousal) of images representing human figures explained these time distortions. This study verified whether abstract paintings of different themes by different artistic movements affect time perception. Twenty paintings with different criteria of movement representation were used: for example, stroboscopic images, photographic blur and action-movement (action painting) which were not tested in the studies of subjective time.

Method

The experiment was approved by the Ethics Committee of the University of São Paulo School of Philosophy, Sciences, and Letters in Ribeirão Preto, Brazil.

Fifteen undergraduate students (8 men; M age = 20.26 yr., SD = 2.12) from the University of São Paulo of Ribeirão Preto were randomly invited to participate. They were untrained in visual arts with normal or corrected-to-normal vision.

The experiment was performed during daylight in an isolated, soundproofed room at the central library of the Ribeirão Preto Campus. Digital photographs of 20 abstract paintings selected according to its pictorial compositions, themes and artistic movements were used (Table 1).

Exposure of stimuli and recording of time estimations were done by the *Wave Surfer* program installed on an HP notebook. The tasks were orally explained to the participants. They were positioned facing the central region of the LG 19" monitor and were asked not to count time. The stimuli were exposed by pressing the "presentation" key and their exposures were finalized after 3 s. At this moment, the monitor was filled with white color indicating that the participant could initiate time estimation. Then, immediately after each time observation the participant reproduced the presentation duration by pressing the "initiate" key. The experienced duration of each stimulus was finalized by pressing the "finished" key (reproduction method). The stimuli were presented randomly to the participants. The *Wave Surfer* program also registered the latency of the participants.

Table 1 – List of paintings presented to participants

Painting	Author	Artistic Movement	Theme
1. *Woman with a guitar*	Georges Braque	Cubism	Person
2. *Girl with Mandolin Fanny Tellier*	Pablo Picasso	Cubism	Person
3. *The Guitar Player*	Pablo Picasso	Cubism	Person
4. *Wounded Bird and Cat*	Pablo Picasso	Cubism	Animal
5. *Portrait of Pablo Picasso*	Juan Gris	Cubism	Person
6. *Nude Descending a Staircase*	Marcel Duchamp	Cubism/Surrealism	Person
7. *The Knife Grinder*	Kazimir Malevich	Constructivism	Person
8. *Abstract Speed*	Giacomo Balla	Futurism	Geometric forms
9. *Dynamism of a Dog on a Leash*	Giacomo Balla	Futurism	Animal
10. *Lines of Movement and Dynamic Succession*	Giacomo Balla	Futurism	Animal
11. *Shape and Noise of Motorcyclist*	Giacomo Balla	Futurism	Geometric forms
12. *The Car Has Passed*	Giacomo Balla	Futurism	Geometric forms
13. *Girl Running on a Balcony*	Giacomo Balla	Futurism	Human body part
14. *The Hand of the Violinist*	Giacomo Balla	Futurism	Human body part
15. *Contemplating*	Paul Klee	Expressionism	Geometric forms
16. *Red Wainscot*	Paul Klee	Expressionism	Geometric forms
17. *The Goldfish*	Paul Klee	Expressionism	Animal
18. *Composition VIII*	Wassily Kandinsky	Expressionism	Geometric forms
19. Number 8, 1949	Jackson Pollock	Abstract Expressionism	Geometric forms
20. Number 14 Grey	Jackson Pollock	Abstract Expressionism	Geometric forms

After the time estimations, the participants' task was to observe the paintings and rate, on a 7-point scale (Likert type), the amount of movement represented, complexity, and about the identification of the figure depicted in the painting: movement, complexity and recognition scales, respectively. They were also asked to rate their level of arousal using the Self-Assessment Manikin scale of 5-point scale (Lang, 1980).

The One Way test (ANOVA) without repeated measures and the Student-Newman-Keuls test for *post hoc* comparisons were used to compare the time estimation, latency, movement, and arousal data for all paintings. One way test analyses were also used to compare the data considering the artistic movements and the themes of paintings separately. Student t-test analyses were conducted comparing the mean values of time ratios of the stimuli with the actual time of exposure of 3 s.

Results

The analyses of variance comparing mean time estimation values of the 20 paintings showed no statistical differences. However, statistical differences for latency [$F(19, 280)=1.66$; $p<.05$], movement [$F(19, 240)=8.28$; $p<.001$] and arousal [$F(19, 240)=3.05$; $p<.001$] variables were found. As the paintings were very different in relation to their pictorial compositions, as was confirmed by movement, recognition and arousal scales, analyses of variance considering the artistic movement and the themes of the paintings conjointly were conduced.

Statistical differences were not verified in the analyses of variance of mean time estimation values for geometrical futurist paintings (8, 11, 12; see Table 1); geometrical expressionist paintings (15, 16, 18, 19, 20); futurist paintings representing human body parts (13, 14); and cubist and futurists painting representing animals (4, 9, 10).

On the other hand, the analyses of time estimation values showed statistical differences between the cubist paintings (1, 2, 3, 5, 6) representing abstract human bodies [$F(4, 70)=2.60$; $p<.05$]. *Pos hoc* comparisons showed that *Woman with a guitar* (1) and *The Guitar Player* (3), more abstract paintings, were estimated shorter than *Nude Descending a Staircase* (6), the painting with stroboscopic representation of a human figure ($p<.05$). The t test analyses showed that *Woman with a guitar* (less movement) was underestimated [$t(28)=-2.05$; $p<.05$] and *Nude Descending a Staircase* was marginally overestimated [$t(28)=1.92$; $p<.06$] in relation to the actual time of painting exposures (3 s).

The movement scale data showed that these paintings were differently scored [$F(4, 60)=4.48$; $p<.001$]: *Nude* painting was scored with a higher movement than the other paintings (all $p<.05$). Arousal levels among these paintings were also different [$F(4, 60)=5.32$; $p<.001$]: there was scored more arousal to *Nude* painting than to *Woman with a guitar*, *Girl with Mandolin Fanny Tellier* and *Portrait of Pablo Picasso* paintings (1, 2, 5; all $p<.01$).

Table 2 – Mean values and standard deviation (parentheses) of time estimation, latency, movement, arousal, recognition and complexity by participants for the 20 cubist paintings

Painting	Time (sec.)	Latency (sec.)	Movement	Arousal	Recognition	Complexity
1	2.66 (0.64) -	1.00 (0.60)	2.23 (1.23)	1.62 (0.63)	1.69 (1.25)**	2.53 (2.10)
2	2.84 (0.70)	1.14 (0.74)	3.00 (1.52)	2.23 (0.92)	6.69 (0.48)	3.38 (1.85)
3	2.88 (1.03)	1.01 (0.70)	3.23 (1.83)	2.69 (0.75)	6.61 (0.76)	2.69 (1.70)
5	3.08 (0.64)	0.99 (0.30)	2.76 (1.23)	2.07 (1.18)	6.84 (0.37)	2.69 (2.17)
6	3.65 (1.30) +	1.88 (1.16)*	4.53 (1.39)*	3.30 (1.25)*	4.30 (1.93)**	4.30 (1.65)

(-) Underestimated; (+) Overestimated; (*) 1 and 3 < 6; (**) 1 and 6 different from 7 points score.

The analyses of variance for the complexity scale did not show statistical differences among the paintings (1, 2, 3, 5, 6). However, for the recognition scale the paintings were differently scored in

dicating that the abstraction of the human figure was not easily evident [$F(4, 60)=52.03$; $p<.001$]. Considering a 7-point score as indicatory of more evident identification of human figure, t-test analyses showed that *Woman with a guitar* ($t(24)=15.30$; $p<.001$) and *Nude Descending a Staircase* [$t(24)=5.02$; $p<.001$] were less easily associated to the human image than were the other three paintings.

The mean latency scores were also statistically different [*F*(4, 70)=3.89; *p*<.01]. This analysis showed that the *Nude Descending a Staircase* painting was different from the other paintings (all *p*<.05). The data for all of the cubist painting variables are presented in Table 2.

Discussion

This work showed that abstract paintings exposed for 3 s affected time perception. However, this effect was associated with the interaction between the painting themes and their artistic movements: only the cubist paintings representing human figures were differently estimated. Stoyanova, Yakimoff, Gourevich, and Mitran (1987) also verified that figurative and abstract paintings presented for short intervals (0.6 to 1.5 s) distorted time perception. They pointed out that the under- and overestimations were caused by the complexity of the artworks but did not describe which their pictorial features were responsible for these results.

The cubist paintings used in this work are mostly brown, ocher, and green. Representing the analytical phase of cubism, they were almost monochromatic analyzing the form without the distraction of the color (Strickland, 2004). On the other hand, they differ in relation to the amount of pictorial abstraction resulting of the different 3-D planes of the human figure representation which causes distinct interpretative ambiguities.

Nicki, Lee, and Moss (1981) showed that subjects observed cubist paintings for more time according to the increase of the subjective ambiguity of paintings. They pointed that interest and pleasure of subjects were associated to the identification of an object or person represented in the pieces. Kuchinke, Trapp, Leder, and Jacobs (2009) asked subjects to indicate the moment when they recognized any object in several cubist paintings. The paintings with high content accessibility were processed faster toward the point of explicit classification of the depicted content. These works showed that affective aesthetic responses depend on the ease with which an artwork can be processed.

In this study, all paintings received scores not different for complexity scale. However, *Woman with a guitar* and *Nude Descending a Staircase* were less easily associated to the human image (recognition scale). Also, these two paintings that were differently scored by the movement scale were under- and overestimated in relation to the actual time of the painting exposures (3 s). This information would explain why the stroboscopic painting *Nude* (more movement) was estimated longer than the other cubist paintings - as *Woman* (less movement) - which do not represent human intentional movements (see Table 2).

Hekker and van Wieringen (1990) used the reaction time to access information about the time required for subjects to recognize human figures in cubist paintings. The results showed that *Nude Descending a Staircase* was recognized after 7.5 s while *Portrait of Pablo Picasso* was quickly identified (less than 1 s). Paintings with similar composition of *Woman with a guitar* were recognized above 9.5 s. The authors pointed that aesthetical pleasure was related to capacity of recognize human figures independently of paintings complexity. Note that *Nude* received the greatest score for latency (Table 2).

Pictorial characteristics of cubist paintings were analyzed by different perspectives involving hedonic factors, pleasure and emotions. Nather et al. (2011) showed that the duration of figurative human figures (Degas' ballerinas) were differently estimated using shorter durations of the stimuli exposures (0.4 to 1.6 s). However, the magnitude of the time distortions was relatively smaller using long exposures (2 to 8 s) due to this the body postures requiring more movement were judged with more arousal. Embodiment mechanisms were associated to these time distortions.

Nude Descending a Staircase painting received the highest arousal score. It is possible to infer that the same mechanisms of time estimations were required when people observed abstract or figurative human figures representing movements of different kinds and intensities. Future researches using artworks will highlight different aspects of time perception related to embodiment mechanisms of movement perception in static images.

Acknowledgements

Francisco C. Nather received a São Paulo Research Foundation (FAPESP) Postdoctoral Scholarship and Young Investigators Awards support. Paola A. M. Fernandes received PIBIC/Santader Undergraduate Research Scholarship. José Lino O. Bueno received a Research Grant from the National Counsel for Technological and Scientific Development (CNPq).

References

Angrili, A., Cherubini, P., Pavese, A., & Manfredini, S. (1997). The influence of affective factors on time perception. *Perception & Psychophysics*, *59*, 972-982.

Braddick, O. (1995). The many faces of motion perception. In Braddick, O., org. *The artful eye* (pp. 205-231). Oxford: Oxford University Press.

Cupchick, G. C., & Gebotys, R. (1988). The experience of time, pleasure, and interest during aesthetic episodes. *Empirical Studies of the Arts*, *6*, 1-12.

Cutting, J. E. (2002). Representing motion in a static image: constraints and parallels in art, science, and popular culture. *Perception*, *31*, 165-1193.

Freedberg, D., & Gallese, V. (2007). Motion, emotion and empathy in esthetic experience. *Trends in Cognitive Sciences*, *11*, 197-201.

Hekker, P.; & van Wieringen, P. C. W. (1990). Complexity and prototypicality as determinants of the appraisal of cubist paintings. *British Journal of Psychology*, *81*, 483-495.

Kim, C. Y.; & Blake, R. (2007). Brain activity accompanying perception of implied motion in abstract paintings. *Spatial Vision*, *20*, 545-560.

Kourtzi Z., & Kanwisher, N. (2000). Activation in human MT/MST by static images with implied motion. *Journal of Cognitive Neuroscience*, *12*, 48-55.

Kuchinke, L.; Trapp, S.; Jacobs, A. M.; & Leder, H. (2009). Pupillary responses in art appreciation: Effects of aesthetic emotions. *Psychology of Aesthetics, Creativity, and the Arts*, *3*, 156-163.

Lang, P. J. (1980) Behavioral treatment and biobehavioral assessment: Computer applications. In Sidowski J., Johnson J., Williams T., eds. *Technology in mental health care delivery systems* (pp. 119-137). Norwood NJ: Ablex.

Nather, F. C., & Bueno, J. L. O. (2011) Static images with different induced intensities of human body movement affect subjective time. *Perceptual and Motor Skills*, *113*, 157-170.

Nather, F. C., Bueno, J. L. O., Bigand, E., & Droit-Volet, S. (2011). Time changes with the embodiment of another's body posture. *PLoS ONE*, *6*, e19818.

Nicki, R. M.; Lee, P. L.; & Moss, V. (1981). Ambiguity, cubist works of art, and preference. *Acta Psychologica*, *49*, 27-41.

Stoyanova, Y., Yakimoff, N.; Gourevich, A.; & and Mitrani, L. (1987). In search of paintings influence on time judgment. *Acta Neurobiologiae Experimentalis*, *47*, 103-109.

Strickland, C. (2004). *The annotated Mona Lisa*. Kansas City: John Boswell Management Inc.

Thakral, P. P.; Moo, L. R.; & Slotnick, S. D. (2012). A neural mechanism for aesthetic experience. *NeuroReport*, *23*, 310-313.

MOTION ILLUSIONS CAUSED BY PAINTINGS OF OP ART DISTORT THE PERCEPTION OF TIME

Francisco Carlos Nather, Fernando Figueiredo Mecca and José Lino Oliveira Bueno
Department of Psychology, University of São Paulo, Faculty of Philosophy Sciences and Letters of Ribeirão Preto, São Paulo, Brazil
f.nather@hotmail.com / jldobuen@ffclrp.usp.br

Abstract

Figurative static images implying human body movements observed for different durations (2-8 and 36 s) affected timing perception. This study examined whether images of static geometric shapes - juxtaposed lines causing visual motion illusions - would affect the perception of time. Undergraduate participants observed two optical paintings by Bridget Riley for 9 or 36 s seconds (G9 and G36 groups). The paintings implying different movements (2.0- and 6.0-point stimuli) were randomly presented. The prospective paradigm in the reproduction method to record the time estimations of the participants was used. Data analysis showed no time distortions in G9 group. In the G36 group the paintings were differently perceived: 2.0-point was estimated shorter than 6.0-point. Also, the exhibition time of the 2.0-point painting was underestimated, compared with the real time. These results show that optical illusion of movement in static images caused time distortions related to a long duration of exposure.

The Optical Art or Op Art can be considered as a generator of perceptual responses, having dynamic qualities which provoke illusions and sensations in the viewer (Rycroft, 2005). Some paintings of this artistic movement use perspective illusions or chromatic tension that lead to the perception of a flicker or movement in simple geometrical patterns (Zanker, 2004).

An important exponent of Op Art is Bridget Riley, whose works has been studied by several experimental studies on optical illusions and movement perception (Dodgson, 2008; Zanker, 2004; Zanker & Walker, 2004; Zeki & Lamb, 1994). According to Rycroft (2005), viewing a Riley painting was intended to be a physical, embodied experience that activated and exercised cognitive consciousness because of the visual movement perception of her paintings, e. g. *Fall* and *Current*, evoke on the viewers (see also Follin, 2004; Moorhouse, 2003).

Fall and *Current* artworks are composed of simple geometric patterns (curved lines) painted in black and white[1]. With the intention of explaining why these painting evoke visual illusions of movement, Zanker (2004) created digitalized images based on the Riley's painting (*Fall*) to identify movement indicators implicit in his images. The movement illusion caused by *Fall* on subjects occurs due to a relationship between these movement indicators and the saccadic eye movements (Zanker & Walker, 2004). Zanker, Hermens, and Walker (2010) found that the factors that intensify these illusion movements are the contrast and the wave frequency of the lines of the painting composition.

Op Art has often been portrayed as an art of high science, a rigorous, retinal art linked to theories of visual illusions and movement perception. However, other artistic movements have *represented* the movement in different ways. *Tension*, force lines, vector arrows, texture, and other technical strategies have been used by artists to represent the

---

[1] Digital reproduction of the painting *Fall* can be viewed on the site http://www.tate.org.uk/artworks/riley-fall-t00616.

movement in abstract paintings, drawings and photography (Arnheim, 2004; Cutting; 2002). According to the authors, the methods are not mutually exclusive and are often combined.

Kim and Blake (2007) used abstract paintings composed of stroboscopic images in a research on the perception of implied movement. They showed that abstract paintings of different artistic movements (Cubism and Futurism) were evaluated with more movement and activated the MT+ brain area more than other paintings that were not intended to evoke movement (Expressionism). Also, subjects with previous experience in abstract arts rated the "stroboscopic" paintings as having more movement than the non-experienced subjects. In a similar way, Riley's paintings were distinctly rated in respect to its evoked movement by untrained subjects in visual arts: *Fall* was rated as having more movement than *Chant 2* in a study that ranked artworks of Riley's from different periods (Giannetti, Nather, & Bueno, 2010)[2].

Similar ranking scales have explored movement perception of different figurative paintings and sculptures. Nather and Bueno (2006, 2008) used photographic reproductions of human-like objects and showed that subjects can recognize movement in these static images relating to different intensities of body movements for each one. Using pictures of sculptures of dancers by Edgar Degas, the authors showed that the perception of movement modulates the subjective time perception. Presented for 36 s, static dancers with lower movement scores were judged to have shorter duration than those with intermediary scores, and the dancers with intermediary scores were judged to be shorter than those with the highest scores (Nather & Bueno, 2011).

Dynamic balance criteria of movement representation could explain why the body postures of these Degas' ballerinas were rated differently. The relationships among the body parts (head, arms and legs) generated different visual asymmetries in artworks causing distinct movement perception (Cutting, 2002). Furthermore, Freedberg and Gallese (2007) hypothesized that an important element of aesthetic episodes consists of the activation of embodied mechanisms encompassing the simulation of actions and corporeal sensation that were being observed.

Nather, Bueno, Bigand and Droit-Volet (2011) showed that the exposure time of artworks is also a variable that affects time perception. Using the same Degas' dancers in a bisection paradigm, they showed that short exposures (0.4 to 1.6 s) of dancers with more or less implicit movement could affect subjective time perception more strongly than long exposures (2 to 8 s). They attributed these results to the embodiment mechanisms as the increased arousal levels were verified in body postures of dancers representing high intensity of movements.

Geometric shapes in motion observed for 6 to 18 s were estimated with longer duration than stationary squares and it was observed that the faster was the movement, the greater was the subjective time distortion (Brown, 1995). Similarly, figurative static images implying motion observed for different durations (0.4 to 1.6 and 36 s) modulated timing perception. The aims of the present study were to verify whether: (a) photographic reproductions of Bridget Riley's paintings rated with different intensities of movement affect time perception; and (b) their different time exposures (9 s and 36 s) differently affect the perception of time.

Method

The experiment was approved by the Ethics Committee of the University of São Paulo School of Philosophy, Sciences and Letters, at Ribeirão Preto, Brazil.

---

[2] *Chant 2* painting can be viewed on the site http://www.op-art.co.uk/op-art-gallery/bridget-riley/chant-2.

Forty-nine university students (23 men; 23.37±4.77 years of age) randomly invited from University of São Paulo, untrained in visual arts with normal or corrected-to-normal vision, participated in the experiment.

The experiment was performed in an isolated, soundproofed room at the central library of the University of São Paulo of Ribeirão Preto Campus. Indirect sunlight was used during the day, and artificial light was used by the night.

Digital photographs of three paintings by Bridget Riley (*Chant 2*, *Fall* and *Exposure*) having respectively 2.0-, 6.0-, and 5.5- points of implied movement were used as stimuli: A, B and Training Stimulus, respectively. As both *Fall* and *Exposure* are black and white paintings, *Chant 2* was modified to a black and white image, so that the colors wouldn't be an additional variable. These stimuli were obtained by using digital reproductions of 19 paintings by Bridget Riley which were scored (Likert 7-point scale) by participants who were untrained in arts.

Exposure of stimuli and recording of time estimations were done by the *Wave Surfer* program installed on an HP notebook. The tasks were orally explained to the participants. They were positioned facing the central region of the LG 19" monitor and were asked not to count time. The stimuli were exposed by pressing the "presentation" key and their time exposure were finalized after 9 s (Group G9) or 36 s (Group G36). At this moment, the monitor was filled with white color indicating that the participant could initiate time estimation. Then, immediately after each time observation the participant reproduced the presentation duration by pressing the "initiate" key. The experienced duration of each stimulus was finalized by pressing the "finished" key (reproduction method). The stimuli were presented randomly in two orders (A-B and B-A) to the participants. The Training Stimulus was presented first to the participants to make them familiar with these experimental tasks. This stimulus data were excluded from the analysis.

After the time estimations, the participants' task was to observe and judge the movement intensity of the paintings answering the Differential Semantic Scales for the locutions "Movement" and "Complexity". Subsequently, they answered questions about the characteristics of the paintings and personal information.

The temporal ratio (reproduced time/real time of exposure) was used in the analysis. The Two Way test (ANOVA) with no repeated measures and the Student-Newman-Keuls test for post-hoc comparisons were conduced individually for the G9 and G36 groups. Student *t* test analyses were conducted comparing the mean values of time ratios of the stimuli with the value 1.0, considering this value as an indication of no time distortion (under or overestimation).

Results

The mean values for temporal estimations (temporal ratio) are shown on Figure 1.

The analyses of variance showed that the order of the stimuli presentation did not affect time estimation of the participants in Group G9 (9 s of images exposure). Also, these analyses did not show differences between stimuli. The *t*-test analysis showed that the temporal estimations of A and B stimuli were not different of the actual duration of 9 s.

The analyses of variance of the G36 Group (36 s of exposure) did not show an effect of the order presentation of images but an effect between stimuli was observed: Stimulus A was estimated shorter than Stimulus B [$F(1, 38)=4.99$; $p<.05$]. The *t*-test analyses showed that the Stimulus A was underestimated in relation to the actual time of exposure of 36 s [$t(20)=-1.96$; $p<.05$].

The majority of participants of both G9 and G36 groups (59.3% and 68.2%, respectively) related that the images caused confusion and/or blurred vision.

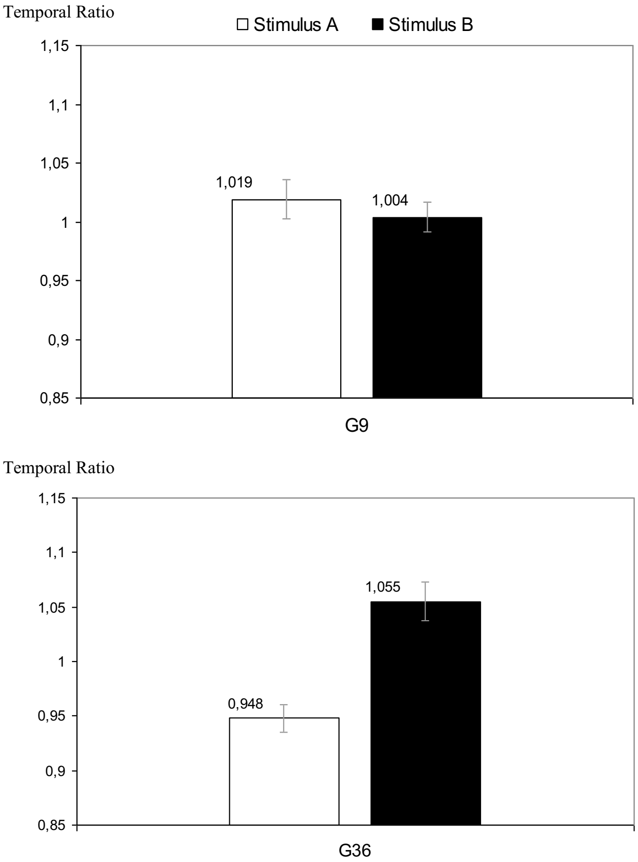


Figure 1: Temporal ratio of the mean values of time estimations: Stimuli A (less movement) and Stimuli B (more movement) in G9 and G36 groups.

Discussion

This study showed that time perception was affected by abstract paintings inducing different intensities of motion illusions. However, this time distortions were observed when the paintings were exposed for 36 s. This result is consistent with previous studies that used figurative images of Degas' sculptures representing different body movements that were presented for the same duration: static dancers with lower movement were judged to have shorter duration than those dancers with the highest movement (Nather & Bueno, 2006, 2011).

On the other hand, time estimations for 9 s of exposure did not cause time distortions between the paintings. This result was not totally in agreement with Nather, Bueno, Bigand, and Droit-Volet (2011) which showed that time distortions between the images were strong for 0.4 to 1.6 s (short) of exposure while the same was not verified for 2 to 8 s (long) of exposure. Further, this result for long durations was not in agreement with those that used stationary and geometric shapes in motion presented for 6 to 18 s (Brown, 1995).

In this sense, the results pointed out that the duration of stimuli is a relevant variable in studies that use static images – artworks representing different abstract and figurative movements – to understand aspects underlying the subjective time perception. Recently, Nather and Bueno (2012) used a behavioral measure allowing the viewer to freely observe (visual exploration) the same images of Degas' dancers and then record this self-observation. They verified that participants observe for a mean of 18 s the different artworks but the time distortions were less marked.

Different methodologies for time estimation with, for example, changes in time exposure and tasks have been used to examine different aspects and processes related to the perception of time (Fraisse, 1984; Block, 1990). From this perspective, it is possible to discuss the different results found between 9 and 36 s of exposure by relating them to the effects of the implicit effects (motion illusions) of Riley's paintings caused on subjects.

According to Rycroft (2005), the aesthetical experience of seeing Riley's paintings such as *Fall* was not corporeally limited to the eye and the mind of the viewer but embodied. It was internalized as a changing set of physical responses and realizations. This can explained why, for example, people related physical states of dizziness and general disorientation when viewing the paintings for an extended time in a museum. In this study the majority of the participants - 60% and 70% (G9 and G36 respectively) - declared that the paintings caused confusion and/or blurred vision.

From this perspective, it is possible to infer that the different duration exposures interact with the effects of motion illusions causing the distinct time distortions in the G36 group. Thereby, not only the visual perception of movement but the body sensations contribute in the process of time perception. According to Rycroft (2005), viewing a Riley painting was intended to be an embodied experience that activated different cognitive processes involved in visual motion experience and eye movements (see Zanker & Walker, 2004; Zanker, Hermens, & Walker, 2010).

The effects of embodiment on these time estimations are also important to be considered. Nather et al. (2011) associated time distortions to embodiment mechanisms as they observed increased arousal levels in short exposures (0.4 to 1.6 s) of images implying human body movements of more intensity. As the arousal levels decayed after 2-3 s of exposure of images with emotional contents (Angrili, Cherubini, Pavese, & Manfredini; 1997), they explain why 2 to 8 s of exposure caused less time distortions. This could explain why 9 s of exposure of Bridget Riley's paintings did not affect time estimations as was observed when these paintings were exposed for 36 s.

Acknowledgements

Francisco C. Nather received a Postdoctoral Scholarship and a Young Investigators Awards support both from São Paulo Research Foundation (FAPESP). Fernando F. Mecca received a CNPq Undergraduate Research Scholarship. José Lino O. Bueno received a Research Grant from the National Counsel for Technological and Scientific Development (CNPq).

References

Angrili, A., Cherubini, P., Pavese, A., & Manfredini, S. (1997). The influence of affective factors on time perception. *Perception & Psychophysics*, *59*, 972-982.

Arnheim, R. (2004). *Art and visual perception: A psychology of the creative eye*. Berkeley: University of California Press.

Block, R. A. (1990). Models of psychological time. In: Block, R. A. (Org.). *Cognitive Models of phychological time* (pp. 1-35). New Jersey: Lea Publishers.

Brown, S. W. (1995). Time, change, and motion: The effects of stimulus movement on temporal perception. *Perception & Psychophysics*, *1*, 105-116.

Cutting, J. E. (2002). Representing motion in a static image: constraints and parallels in art, science, and popular culture. *Perception*, *31*, 165-1193.

Dodgson, N. A. (2008). Regularity and randomness in Bridget Riley's early Op Art. In *Proceedings of Eurographics Workshop on Computational Aesthetics in Graphics, Visualization, and Imaging, Eurographic Association*, 107-114.

Follin, F. (2004). *Embodied visions: Bridget Riley, Op Art and the sixties*. London: Thames and Hudson.

Fraisse, P. (1984). Perception and estimation of time. *Annual Review of Psychology*, *35*, 1-36.

Freedberg, D., & Gallese, V. (2007). Motion, emotion and empathy in esthetic experience. *Trends in Cognitive Sciences*, *11*, 197-201.

Giannetti, S. A. L., Nather, F. C., & Bueno, J. L. O. (2010). Percepção de movimento em obras de arte abstrata e seu efeito na percepção subjetiva de tempo. In *18º Simpósio Internacional de Iniciação Científica*.

Moorehouse, P. (2003). *Bridget Riley*. London: Tate Gallery Publishing.

Kim, C. Y.; & Blake, R. (2007). Brain activity accompanying perception of implied motion in abstract paintings. *Spatial Vision*, *20*, 545-560.

Nather, F. C., & Bueno, J. L. O. (2006). Efeito de imagens estáticas com diferentes representações de movimento sobre a percepção subjetiva de tempo. *Psicologia: Reflexão & Crítica*, *19*, 217-224.

Nather, F. C., & Bueno, J. L. O. (2008). Movement ranking scale of human body static images for subjective timing estimation. In *Proceedings of the Annual Meeting of the Society for Psychophysics*, *24*, 185-190.

Nather, F. C., & Bueno, J. L. O. (2011). Static images with different induced intensities of human body movements affect subjective time. *Perceptual and Motor Skills*, *113*, 157-170.

Nather, F. C., Bueno, J. L. O., Bigand, E., & Droit-Volet, S. (2011). Time changes with the embodiment of another's body posture. *PLoS ONE*, *6*, 19818.

Nather, F. C., Bueno, J. L. O. (2012). Exploration time of static images implying different body movements causes time distortions. *Perceptual and Motor Skills*, *115*, 105-110.

Rycroft, S. (2005). The nature of Op Art: Bridget Riley and the art of nonrepresentation. *Environment and Planning D: Society and Space*, *23*, 351 – 371.

Zanker, J. M. (2004). Looking at Op Art from a computational viewpoint. *Spatial Vision*, *17*, 75-94.

Zanker, J. M.; & Walker, R. (2004). A new look at Op art: towards a simple explanation of illusory motion. *Naturwissenchaft*, *91*, 149-156.

Zanker, J. M.; Hermens, F.; & Walker, R. (2010). Quantifying and modeling the strength of motion illusions perceived in static patterns. *Journal of Vision*, *10*, 1-14.

Zeki. S.; & Lamb. M. (1994). The neurology of Kinectic Art. *Brain*, *117*, 607-636.

The Identification of Non-Native Speech Sounds with Psychophysical Training

Jordan Richard Schoenherr, John Logan, and Cassandra Larose
psychophysics.lab@gmail.com, john_logan@carleton.ca, clarose@connect.carleton.ca
Department of Psychology, Carleton University
1125 Colonel By Drive, Ottawa, ON K1S5B6 Canada

Abstract

Many early studies of speech perception assumed that the psychophysical properties of speech sounds were unavailable during stimulus identification. When participants are presented stimuli along an acoustic continuum and perform a 2AFC, identification performance is well-described by a logistic function suggesting two discrete phonemic categories. These phonemic category boundaries are believed to bias the classification of speech sounds from non-native languages, reducing the ability to detect acoustic differences. Using a brief period of psychophysical training and a phoneme identification task, participants were sensitized to differences along a voice-onset time (/b-p/) continuum enabling the identification of a non-native phoneme (/p^h/). Finally, participant confidence reports suggested that they were generally unaware of their capacity to accomplish the task. Results demonstrate that participants could use psychophysical properties of the stimuli to identify non-native speech sounds while subjective confidence reports indicated that they had varying degrees of awareness of the psychophysical properties of the non-native speech sounds.

The acquisition of non-native phonemic categories by listeners can be difficult due to interference from prior linguistic experience. Acoustic differences between speech sounds that listeners might otherwise be capable of detecting become undifferentiated due to the development of native language phoneme categories within the first year of life (Werker, 1989) with some suggesting that after infancy, adults lose the ability to discriminate these sounds (Eimas, 1975). For instance, along the voice-onset time (VOT) continuum, English listeners learn to only differentiate /b/ and /p/ based on the difference between the acoustic cues associated with aspiration and the vibration of the vocal cords. When presented with stimuli from the non-native phoneme category portion of this continuum (VOT < -30), corresponding to the Thai /p^h/ category, adult monolingual English listeners would not be capable of distinguishing these phonemes from the neighbouring /b/ speech sounds. Results such as these were initially taken as suggesting that adult listeners no longer had access to these acoustic properties (e.g., Liberman, Harris, Hoffman, & Griffith, 1957). However, this conclusion was later challenged by findings that response times varied with acoustic differences within categories (Pisoni, 1973; Pisoni & Tash, 1974) suggesting that the amount of evidence that needed to be accumulated was dependent on both the phonemic and acoustic properties of the stimuli. In the present study, a psychophysical training technique developed by Pisoni, Aslin, Perey, and Hennessy (1982) was used to assess listeners' awareness of acoustic differences during the development of the non-native speech category /p^h/.

In an ecologically valid setting of language learning, listeners are presented with a number of natural speech tokens within their linguistic environment. One extensively studied example of this is the acquisition of the /r-l/ liquid distinction by Japanese listeners. Specifically, MacKain, Best, and Strange (1981) required listeners to identify and discriminate (in AXB and oddity tasks) synthetic speech sounds along this continuum, American listeners produced sharper category boundaries and more accurate discrimination performance than Japanese listeners with or without 'intensive training'. Studies using natural tokens produced by native and non-native speakers also produce similarly poor performance

(Sheldon & Strange, 1982). However, training techniques have proven effective in training Japanese listeners by increasing stimulus variability and task demands (Logan, Lively, & Pisoni, 1991).

Attention is required in order to learn to discriminate and identify stimuli along a perceptual dimension (Nosofsky, 1986). Jusczyk (1992) argued that portions of an acoustic continuum that are typically left unattended can be accurately discriminated only when selective attention is allocated to those physical dimensions. Evidence in support of these claims comes from psychophysical training techniques developed by Pisoni, et al. (1982) wherein native English listeners were presented with three exemplar speech sounds from regions along the VOT continuum corresponding to voiceless unaspirated, voiced aspirated, and voiceless aspirated stops (i.e, /p/, /b/, and /p^h/, respectively). Listeners were asked to identify each speech sound and received feedback over multiple days of training. In contrast to earlier studies (e.g., MacKain et al., 1981), Pisoni et al. (1982) observed that listeners could discriminate the non-native /p^h/ category from the neighbouring /b/ category with little to no confusion with the /p/ category. These findings suggest that the focus of selective attention can facilitate acquisition of non-native phonemes and that participants can develop this new category of speech sounds in contrast to the neighbouring category (i.e., /b/). An open question that remains, however, is whether listeners have an awareness of the properties of this new phoneme category or whether instead the learning that has occurred is below the threshold of subjective awareness.

Participants' subjective awareness is assessed by having them provide estimates of the subjective probability that they provided the correct answer. When using numerical scales in a two-alternative forced choice (2AFC) task, 50% represents a response associated with a guess whereas 100% represents absolute certainty in a response. Those listeners that assign their subjective probabilities appropriately to their responses (e.g., p(cor) = 0.5 and mean(conf.) = 50%) are said to be perfectly calibrated. In contrast to this ideal, participants in these experiments are frequently miscalibrated. In perceptual tasks, underconfidence is typically observed with accuracy exceeding confidence (e.g., Bjorkman, Juslin, & Winman, 1993) whereas in tasks assessing general knowledge, overconfidence is typically observed with confidence exceeding accuracy (e.g., Gigerenzer, Hoffrage, & Kleinbolting, 1991). These findings have been explained by suggesting that underconfidence might result from a lack of awareness of one's perceptual system (Dawes, 1980) or that task difficulty produces differences in subjective calibration (Lichtenstein & Fischhoff, 1977). Overconfidence has also been observed in perceptual tasks. Using a perceptual discrimination task, Baranski and Petrusic (1994) observed overconfidence when participants were presented with pairs of line-lengths that were difficult to discriminate. Such findings also provide evidence against single-process accounts of confidence based solely on the primary decision processes (e.g., Ferrel & McGooey, 1980) while suggesting that additional processes are required to compute a confidence report (for a review, see Baranski & Petrusic, 1998).

The formation of non-native phoneme categories by adult listeners suggests that listeners maintain a capacity to perceive acoustic information from a speech signal (cf. Eimas, 1975) even though their proclivity is to not attend to such information (Werker, 1989). In the present study, we examine whether listeners possess an awareness of these acoustic differences by having them rate their subjective confidence. For the purposes of this study we assume that the information accumulated for one phoneme category (e.g, /p^h/) is contrasted against that accumulated for the neighbouring category along the VOT continuum (e.g., /b/). Thus, when presented with a stimulus the probability of a guess represents 50% thereby allowing the use of a standard 6-point scale from guess (50%) to certainty (100%). If participants are aware of acoustic differences, their subjective confidence should differ across regions of the VOT continuum as evidenced by miscalibration. If underconfidence is evidenced, it suggests that listeners did not have subjective awareness of a well-defined phonemic category whereas if overconfidence is evidenced, it suggests that listeners believed that they had a better understanding of the phonemic category then they in fact did.

Method

Nine Carleton University students participated in the study for course credit; all were native speakers of English or had extensive experience with English and reported normal hearing and no speech pathologies. Fifteen synthetic speech stimuli were used, obtained from the Haskins Laboratories website (HL, 2011; Lisker & Abramson, 1967). These stimuli varied along the VOT continuum from -70 to 70 ms VOT. As per the method used by Pisoni et al. (1982), listeners were presented with stimuli which corresponded to the prevoiced phoneme category /p^h/ for those stimuli in the negative VOT range, and the /b/ and /p/ phoneme categories for the remainder of the range. The latter categories are present in English while the former is not. The sounds were originally recorded on reel-to-reel tape and later converted into AIFF format at Haskins Laboratories. Stimuli were pre-processed using a DC offset correction to eliminate clicks present in the AIFF versions and then converted into WAV files.

Procedure

Modelled after Pisoni et al. (1982), listeners were presented with a brief training block in which they were presented with three stimuli prior to the identification tasks, one from each region of the VOT continuum (-70, 0, and 70 ms VOT, corresponding to the /p^h/, /b/, and /p/ categories). Ten replications of these stimuli were presented in the order indicated. In the following training identification (ID) task, listeners were provided with feedback. They were presented with a stimulus and then reported whether it was a /p^h/, /b/, or /p/ using the ‘V’, ‘B’, or ‘N’ keys, labeled as ‘_B’, ‘B’, and ‘P’, respectively. After they had indicated their response, ‘Correct’ or ‘Incorrect’ was presented visually on the screen as per the response accuracy. Listeners completed a total of 80 trials in the training task.

In the following ID task, listeners again identified the stimulus presented as a /p^h/, /b/, or /p/ using the keyboard. In the first block, after they completed each ID trial they also indicated their level of confidence in their response using the ‘E’ through ‘I’ keys, on a 6-point scale with 50% representing a guess and 100% representing certainty. In the second block, confidence was not reported. Each block was composed of a total of 150 trials.

The duration of the experiment was approximately 30 minutes. Listeners were presented with the stimuli over headphones using PsychoPy software (Peirce, 2007).

Results

Proportion Identification. Unlike studies that have examined 2 category identification performance using confidence reports (e.g., Schoenherr, Logan, & Larose, 2012), only the /p/ phoneme category showed a sharp identification function (Figures 1a and 1b). In general, however, listeners could consistently identify stimuli associated with the /p^h/ and /b/ category with greater than chance accuracy (i.e., stimuli with VOTs of -70, -60, -50, 0, and 10) indicating that even with a brief period of psychophysical training, listeners can begin to acquire a non-native speech category. Supporting this, we obtained a significant effect for VOT stimulus, $F(14,112) = 7.389$, $MSE = .435$, $p = .001$, $\eta^2 = .480$. Given that we did not obtain a main effect or interaction of confidence reports, it suggests that confidence reports did not significantly affect ID performance thereby permitting a straightforward interpretation of the remaining results.

Figure 1a. Identification Function without the Requirement of Confidence

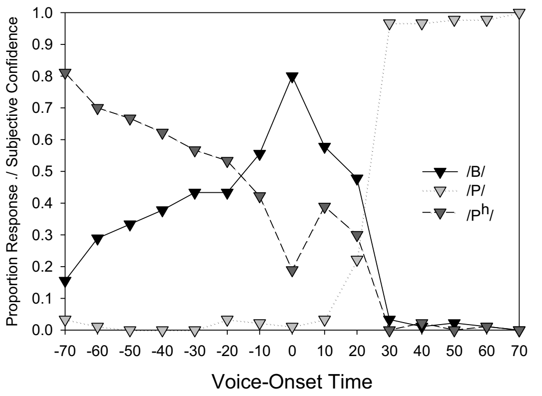
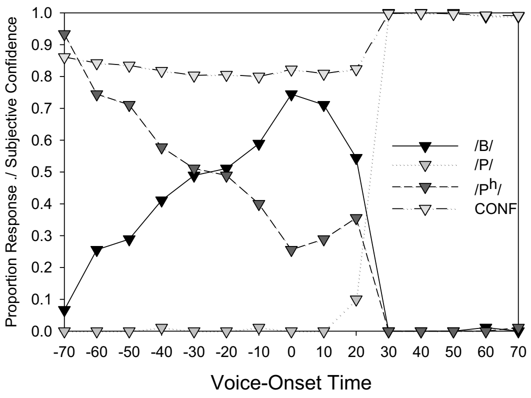


Figure 1b. Identification Function with the Requirement of Confidence



Identification Response Time. Prior to conducting an analysis of the response time data, we collapsed stimuli into regions five regions along the VOT continuum corresponding to two category boundaries (CBs) /p^h-b/ and /b-p/ corresponding to CB$_1$ (-30, -20) and CB$_2$ (20, 30), respectively, and equivalent within-category pairs corresponding to /p^h/ (-70, -60), /b/ (0,10), and /p/ (60, 70), respectively. Using the criterion of 3 standard deviations, 4.3% of the responses were identified as outliers and removed from the final analysis.

Table 1. Mean identification response time in ms along the critical regions of the VOT continuum with standard error reported in parentheses.

Confidence Condition	**/p^h/ (-70, -60)**	**/p^h-b/ (-30, -20)**	**/b/ (0, 10)**	**/b-p/ (20, 30)**	**/p/ (60, 70)**
No Confidence	888 (45)	910 (51)	893 (54)	933 (75)	796 (43)
Confidence	1,009 (47)	1,137 (124)	1,072 (113)	992 (83)	855 (41)

An analysis of the remaining responses times revealed a main effect of VOT region, *F*(4,32) = 4.45, *MSE* = .041, *p* = .025, η^2 = .357. Table 1 indicates that response latencies were longer at category boundaries as well as for the non-native (/p^h/) and modified native (/b/) categories relative to the native /p/ category. A main effect of the requirement of confidence report was also obtained, *F*(1,18) = 14.55, *MSE* = .026, *p* = .005, η^2 = .645. Again, Table 1 demonstrates longer latencies with the requirement of confidence relative to the no confidence condition. Given that the confidence block always followed the no confidence block, this finding cannot be attributed to automaticity. The interaction of confidence condition and VOT region was only marginally significant, *F*(4,32) = 2.724, *MSE* = .019, *p* = .099, η^2 = .254.

Confidence Reports. Figure 1a and 1b also demonstrate the effect of confidence measures. Listeners expressed less confidence in their responses to stimuli located within the /p^h/ and /b/ categories. As was the case with ID accuracy, we observed a main effect of the stimulus location along the VOT continuum on mean confidence, *F*(14,112) = 6.931, *MSE* = 1011.371, *p* = .018, η^2 = .464. Our comparison of over/underconfidence bias did not reveal any significant effects, $F(14,112) = 2.146$, $MSE = .0354$, $p = .133$, $\eta^2 = .212$. Although it is possible that our small sample size might obscure a significant effect due to the inherent individual differences associated with confidence reports, our findings suggest that listeners might not be fully aware of the processes allowing them to identify stimuli.

Discussion

The results of the present study replicated findings obtained by Pisoni, et al. (1982) and Baranksi and Petrusic (1994, 1998). With minimal psychophysical training, we were able to induce listeners to perceive a non-native speech category in the voiceless, unaspirated portion of the VOT continuum (i.e., /p^h/) (Pisoni et al., 1982). Importantly, however, listeners' identification functions were not as sharp as those obtained by Pisoni et al. (1982): listeners' performance was lower for stimuli within the /b/ category. One possibility is that the stimuli used in the present study (HL, 2011) might not have provided appropriate acoustic cues to allow participants to acquire the non-native speech category. Alternatively, and more plausible, the reduced quantity of training provided in the present study might have resulted in phonemes that were not as well-defined to the listeners. This provides further support for claims that participants cannot only discriminate acoustic differences within a category boundary (e.g., Iverson & Kuhl, 1995; Miller & Volatis, 1989; Pisoni 1973; Pisoni et al., 1982; Pisoni & Tash, 1974) but that phoneme categories are somewhat plastic in adult listeners (cf. Eimas, 1975; Werker, 1989). Consequently, we can proceed to interpret confidence ratings directly.

When asked to report confidence, listeners took additional time to perform the primary decision. This additional requirement did not however affect performance. Listeners' accuracy in identifying stimuli did not significantly differ between confidence and no confidence conditions, even though the block in which listeners reported confidence always followed the no confidence condition leading to the possibility of training effects. An examination of mean confidence revealed that listeners expressed less confidence in the /p^h-b/ portion of the VOT continuum, suggesting that they had less certainty in their responses. Moreover, the overconfidence expressed by listeners across that portion of the continuum suggests that even if the acoustic properties of the auditory signal were available to listeners, they relied on phonemic representations when reporting the level of certainty in their response.

An instructive comparison can also be made between the results of the present study and those of a comparable 2AFC variant of the task: Schoenherr et al. (2012) observed a small decrease in confidence around the /b-p/ category boundary when only the voiced and voiceless portions of the continuum were presented to listeners. When the prevoiced portion of the continuum was additionally presented, Schoenherr et al. (2012) observed lower confidence in the /p^h-b/ portion of the continuum. When compared to identification accuracy, this pattern of responses leads to underconfidence in comparison to the overconfidence observed in the present study. Taken together with our results, this suggests that training does result in the allocation of attention to newly relevant acoustic properties of the stimuli in this region of the VOT continuum, thereby reducing certainty. Although listeners are somewhat more conservative in their confidence reports, they are still overconfident suggesting that the phonemic representations that they are subjectively aware of are less accurate than the acoustic information necessary to identify the stimuli.

References

Abramson, A., & Lisker, L. (1965). Voice onset time in stop consonants: Acoustic analysis and synthesis. Proceedings *of the Fifth International Congress on Acoustics*, Liege, A51.

Baranski, J. V., & Petrusic, W. M. (1994). The calibration and resolution of confidence in perceptual judgements. *Perception & Psychophysics*, *55*, 412-428.

Baranski, J. V., & Petrusic, W. M. (1998). Probing the locus of confidence judgments: Experiments on the time to determine confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 929-945.

Eimas, P. D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the [r-1] distinction by young infants. *Perception & Psychophysics, 18,* 341-347.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *98*, 506-528.

Haskins Laboratories (2011). Abramson/Lisker VOT Stimuli. Retrieved 01/12/2011. From http://www.haskins.yale.edu/featured/demo-liskabram/index.html /.

Jusczyk, P. (1989). Developing phonological categories from the speech signal. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological Development: Models, Research, Implications*. Monkton, MD: York.

Kvidera, S., & Koustaal, W. (2008). Confidence and decision type under matched stimulus conditions: overconfidence in perceptual but not conceptual *Decisions. Journal of Behavioral Decision Making*, *21*, 253–281.

Liberman, A.M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*, 358-368.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how, much they know? *Organizational Behavior and Human Performance*, 20, 159-183.

Lisker, L, & Abramson, A. S. (1967). The voicing dimension: Some experiments in comparative phonetics. *Proceedings of the 6th International Congress of Phonetic Sciences.* Prague: Academia.

Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, *46*, 505-512.

Nosofsky. R. (1986). Attention, similarity, and the identification-categorization relationship. Journal of Experimental Psychology: General, 115, 39-57.

Peirce, J. W. (2007) PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*, 8-13.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics, 13,* 253-260.

Pisoni, D. B., Aslin, R. N., Percy, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 297-314.

Pisoni, D. B., & Tash, J. B. (1974) Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics, 15,* 285-290.

Schoenherr, J. R., Logan, J., and Larose, C. (2012). Subjective confidence of acoustic and phonemic representations during speech perception. *Proceedings of the 28th Annual Meeting of the International Society for Psychophysics*, Ottawa, Ontario, Canada.

Sheldon, A. & Strange, W. (1982). The acquisition of /r/ and /1/ by Japanese listeners or English: Evidence that speech production can precede speech perception. *Applied Psycholingusitics*, 243-261.

Vickers, D., & Packer, J. S. (1982). Effects of alternating set for speed or accuracy on response time, accuracy, and confidence in a unidimensional discrimination task. *Acta Psychologica*, 50, 179-197.

Werker, J. F. & Logan, J. S. (1985). Cross-language evidence for three-factors in speech perception. *Perception & Psychophysics*, *37*. 35-44.

EFFORTFUL PROCESSING IN THE SPEED-ACCURACY TRADEOFF PHENOMENON

Matthew S. Runge & Craig Leth-Steensen
Department of Psychology, Carleton University, Ottawa, K1G 5B6, Canada
mrunge@connect.carleton.ca, craig_leth_steensen@carleton.ca

Abstract

Forty students participated in a dual-task speed-accuracy trade-off experiment. The primary task required them to compare the relative sizes of two symbolic stimuli, and the secondary task required them to periodically identify a tone which sounded at the beginning of each trial. The primary task was performed under both speed-emphasized and accuracy-emphasized conditions. Measures of response time and proportion correct were taken for both tasks under both conditions. Results for the primary task suggested that participants adhered to the instructional emphases of speed and accuracy. Decremented performance in the secondary task was not evident with respect to proportion correct; however, secondary task response times in the speed-emphasized condition were slower than in the accuracy-emphasized condition. These data provide some evidence that speeded responding might indeed be an inherently effortful process and, hence, might not simply involve passive adjustments to response criteria.

The speed-accuracy trade-off is an important phenomenon in the study of both psychophysics as well as human information processing more generally. Whereas in the latter field this trade-off is often regarded as a potential nuisance factor with respect to interpreting both response time and accuracy data, in psychophysics the speed-accuracy trade-off is regarded as a phenomenon worthy of study in its own right. Indeed, asking participants to trade off speed for accuracy is viewed as a means through which the potential mechanisms of psychophysical decision making can be illuminated (Link, 1992; Petrusic, 1992; Vickers, 1979).

Two key theoretical frameworks within which to understand psychophysical decision making are the sequential-sampling-based random walk (Link, 1992; Ratcliff & Rouder, 1998) and accumulator (Vickers, 1979; Usher & McClelland, 2001) models. Each of these models assumes that the perceptual evidence for any decision is accrued over time (either in small packets or continuously) until a criterion amount of information has been obtained, at which point the response associated with that criterion is evoked. Random walk models regard this process as involving the diffusion of an evidence signal between two (positive and negative) absorbing response-threshold boundaries. Accumulator models regard this process as a buildup of evidence signals within two (or more) accumulators with set decision bounds.

A key assumption of both of these decision frameworks that allows them to account for the presence of speed-accuracy trade-offs surrounds the notion that participants are able to strategically lower their response boundaries under speeded conditions in order to provide faster decision times. Because the evidence signal in both of these decision frameworks is assumed to be noisy rather than deterministic, a lowering of the response boundaries will necessarily lead to more decisional errors (and, hence, lower accuracy). In other words, the lower an incorrect decision boundary is, the greater is the probability that it will be crossed in error. (Although for a completely alternative view of this phenomenon see the parallel sophisticated-guessing model of Meyer, Irwin, Osman, & Kounios, 1988.)

The issue of concern in the present study is the degree to which the lowering of the response boundaries in such models can be regarded as being either a fairly passive process or an active one which might actually require a great deal of effort. In order to measure the effort associated with performing decisions under either speed or accuracy sets, a dual-task methodology will be employed here. When participants are asked to perform two (primary and secondary) tasks simultaneously, any decrements in the performance of either or both tasks as compared to when they are being performed alone can be regarded as indicating that the amount of shared cognitive resources required to perform both tasks at the same time exceeds the available limits (Brown & Merchant, 2007; Kahneman, 1973; Wickens, 1984). Furthermore, decrements in secondary task performance across changes in the nature of the primary task can be regarded as indicating that the resource requirements of primary task have increased, hence, leaving less resources available to perform the secondary task. In the present study, the primary task will involve making size comparisons (under either speed or accuracy set), while the secondary task will involve identifying tones.

Some insight into the potential resource requirements of speeded responding can be found in a recent study by Van Veen, Krug, and Carter (2008) who investigated the neurological basis for speed-accuracy trade-offs through functional magnetic resonance imaging (for some related experimental work see Kleinsorge, 2001). Van Veen et al.'s methodology required participants to respond to a set of stimuli according to a conventional speed-accuracy procedure. Participants were instructed to emphasize either accuracy or speed at the beginning of each mini-block, which itself consisted of 4, 8, or 20 trials. Each trial consisted of a Simon task, wherein participants to respond to the colour of a stimulus appearing either to the left or the right of a fixation point. Imaging data revealed an increase in sustained metabolic baseline activity in the left dorsal premotor cortex, the left intraparietal lobe, basal ganglia/thalamus, and the dorsolateral prefrontal cortex during speed emphasis. Indeed, within all neurological areas measured, greater activation was consistently observed in conditions emphasizing speed. Accuracy-emphasized conditions on the other hand, were consistently accompanied by greater transient activation as indexed by further increases in brain activation (above the baseline level) that occurred in response to the presentation of the Simon stimuli.

The theoretical implications of these imaging results are extensive. Namely, it seems that speed-accuracy trade-offs might be directly influenced by activation of the dorsolateral prefrontal cortex. This area is thought to then increase baseline activation in subsequent decision-making, response-preparation, and motor execution areas. Hence, the results of Van Veen et al. (2008) could be regarded as providing some key insight into the workings of sequential-sampling-based decision models. Namely, that speed-emphasized responding is indeed characterized by a reduction in the distance between the baseline evidential starting points of these models and the response boundary thresholds. However, in accordance with Van Veen et al., this reduction is more likely to be due to active increases in the levels of non-specific baseline activation, as opposed to passive decrease in the levels of the response thresholds (see Bogacz, Wagenmakers, Forstmann, & Nieuwenhuis, 2010, for further discussion of this point).

Method

Participants. Forty Carleton University students (13 males and 27 females) participated in this study. All of them participated for experimental credit towards a first-year psychology course. Each participant performed the task individually, and each session took approximately 45 min.

Stimuli and Apparatus. The name stimuli for the primary comparison task consisted of two separate six-item lists of both animal and inanimate object names taken from the normative size rating results given in the study of Shoben, Čech, Schwanenflugel, and Sailor (1989). These norms were used to equate the rated sizes of the items at each of six ordinal positions across the two lists and also to equate (as best as possible) the interval-level, rated size differences between all pairs of ordinally adjacent items within each list. The mean size ratings for each ordinal list position were -4.80, -2.78, -1.05, 1.21, 3.13, and 4.90 (on a scale of -9 to 9). The two lists were (a) Peach, Book, Chicken, Rake, Bicycle, and Sofa, and (b) Comb, Dove, Pail, Banjo, Table, and Cannon. The stimuli for the secondary tone identification task were three tones with frequencies of 300, 900, and 1500 Hz, respectively. These tones were played through a speaker that was located in the tower of a 486 desktop computer sitting right beside the participants.

The whole experiment was programmed and ran using Micro Experimental Laboratory (MEL V.2.0) on a 486 computer. During each comparison trial, two of the six name stimuli were presented side by side in the center of the computer screen. The font size for the stimuli (as determined by MEL System48 font) was somewhat larger than a regular DOS font size. On all trials, the instruction for comparison task (i.e., "Larger?" or "Smaller?") was first presented at the top of the screen. Comparison task responses were made by pressing either the "z" or "/" keys on the left and right sides of the bottom row of the computer keyboard. Tone task responses were made by pressing one of the "v", "b" or "n" keys in the middle of the bottom row of the computer keyboard. All five keys were marked with colored stickers.

Procedure. Throughout the experiment, participants were seated in front of a computer screen in a dimly lit room. Each trial involved the following series of events. First, one of the three tones sounded for 250 ms. Next, the comparative instruction appeared for 1000 ms along with a temporary plus-sign fixation point (located in the middle of the screen). Then, a pair of name stimuli appeared and participants had to choose which one (i.e., the thing on the left or on the right) was either the smaller or the larger by pressing the appropriate response key with the index fingers of their left or right hand. Both the instruction and the stimulus pair remained on the screen until the response was made. On 25% of the comparison trials, a prompt then appeared asking the participant to report which tone had occurred on that trial (i.e., "the last tone" that they had heard). Participants reported that it was either the low, medium, or high tone by pressing either the left-most, middle, or right-most tone-task response keys, respectively. After 1000 ms, the next trial was initiated.

In each half of the experiment, participants first performed a practice block of 20 trials that was followed by two test blocks of 60 trials. In each respective half, participants were asked either to try to be as accurate as possible for the comparison task or to try to respond as quickly as possible for the comparison task. The order of usage of these two speed-accuracy emphases was completely counterbalanced across all participants. They were also told to always try to remember the tones as best as they could, regardless of the speed or accuracy emphasis placed on the comparison task. When accuracy was emphasized, feedback was provided throughout about the correctness of each comparison task response. When speed was emphasized, a response time deadline of 550 ms was invoked such that participants were given a "Too Slow" message whenever they did not respond before that much time had elapsed after the presentation of the comparison stimuli

Only pairs of stimuli that were adjacent in the ordering were used in the practice block, whereas in the test blocks, all possible pairs of stimuli were used. The order of presentation of both the comparison pairs (in each possible left-right positional ordering) and the smaller or larger instructions was completely randomized within each block of trials. The stimulus set was switched for each half of the experiment. The order of usage of each of the two stimulus sets was counterbalanced across all participants and also completely crossed with the speed-accuracy emphasis conditions. Two random orderings of tone stimuli across trials were derived and used for every participant in each of the two test blocks within each of the accuracy- and speed-emphasized halves of the experimental session (with a part of one of these ordering used for the practice trials). At set places within each of these two orderings (i.e., Trials 5, 10, 12, 16, 19, 25, 28, 29 32, 37, 42, 45, 50, 52, and 57 for the first test block, and Trials 7, 9, 13, 16, 19, 20, 22, 28, 33, 37, 39, 46, 49, 55, and 58 for the second test block) participants were cued to recall the tones. Each ordering contained an equal number of presentations of each of the three tones and, hence, recall for each of the three tones was tested equally.

Results

Only the data obtained from the test blocks were used in the following analyses. Separate analyses were performed for each of the following four dependent variables computed separately for each participant (a) median response times for the primary comparison task across the 120 trials in each of the speed- and accuracy-emphasized conditions, (b) the proportion of correct responses made to the primary comparison task across the 120 trials in each of the speed- and accuracy-emphasized conditions, (c) median response times to the secondary tone task across the 30 tone-recall trials in each of the speed- and accuracy-emphasized conditions, and (d) the proportion of correct responses made to the secondary tone task across the 30 tone-recall trials in each of the speed- and accuracy-emphasized conditions. For each dependent variable, a paired samples *t*-test was performed that contrasted performance under speed and accuracy emphasis (of course, much more extensive analyses of these data could have been performed but such analyses were not deemed necessary for the present purposes). Response time medians were used in order to minimize the effects of any really long response times (although note that it is the means of these medians across the 40 participants that are reported descriptively) and analyses of the proportion correct measures were conducted using arcsine-transformed proportions (even though it is the means for the untransformed accuracy measures that are reported descriptively).

With respect to response times, the *t*-tests revealed that there was a significant difference in response times for the accuracy-emphasized and speed-emphasized conditions of the primary symbolic comparison task; $t(39) = 9.409$, $p < .001$. There was also a significant difference in the response times for the secondary tone identification task across the accuracy-emphasized and speed-emphasized primary task conditions; $t(39) = -2.052$, $p < .05$. The corresponding mean response time results are summarized in Table 1.

Table 1. Mean response times for the speed- and accuracy-emphasized conditions.

	Speed-Emphasis (ms)	Accuracy-Emphasis (ms)	Probability
Symbolic Comparison	$M = 735$, $SD = 278$	$M = 1670$, $SD = 625$	$P < .001$
Tone Identification	$M = 1696$, $SD = 357$	$M = 1581$, $SD = 488$	$P < .05$

With respect to accuracy, the *t*-tests revealed that there was a significant difference in the proportion of correct responses in the accuracy-emphasized and speed-emphasized conditions of the primary symbolic comparison task; $t(39) = 12.658$, $p < .001$. On the other hand, there was no significant difference in the proportion of correct responses for the secondary tone identification task across the accuracy-emphasized and speed-emphasized primary task conditions; $t(39) =.594$, $p < .556$. The corresponding mean proportion correct results are summarized in Table 2.

Table 2. Proportion of correct responses for speed- and accuracy-emphasized conditions.

	Speed-Emphasis	Accuracy-Emphasis	Probability
SymbolicComparison	$M = .714$, $SD = .197$	$M = .933$, $SD = .186$	$P < .001$
Tone Identification	$M = .747$, $SD = .362$	$M = .761$, $SD = .350$	$P < .556$

Discussion

The present experiment utilized a dual-task methodology to determine the cognitive effort affiliated with making speeded responses compared to accurate ones. Methodologically, participants were required to compare the sizes of pairs of symbolic stimuli while intermittently responding to a tonal presentation (i.e., the primary and secondary tasks, respectively). The present results showed that correct responding in the primary task was significantly lower under speed emphasis than under accuracy emphasis indicating that participants adhered to the primary task requirements. However, accuracy for the secondary task was not differentially affected (to any great extent) by the speed-accuracy emphasis placed on the primary task. Moreover, participant response times for the primary task in the speeded condition were significantly faster than in the accuracy condition, also indicating that participants adhered to the primary task requirements. However, for the secondary task, participants were significantly slower when speed was emphasized in the primary task in comparison to when accuracy was emphasized.

It had initially been expected that any increased demands on centrally-located cognitive processing resources that were associated with enhancing baseline levels of neural activation, in order to provide speeded responses to the primary comparison task, should indeed have served to significantly affect the accuracy of responding to the secondary tone identification task. Namely, accurate responding to this secondary task could be regarded as depending on the degree to which the participants could faithfully maintain the identity of the presented tone in memory while responding to the comparison task. Hence, if more resources were required to respond to the comparison task under speed emphasis, less would have been available to allocate to the maintenance of the tone, making it less likely that participants would then have been able to keep track of it. However, it is possible that the nature of the cognitive resources used to invoke speeded comparison responses and those used to maintain the memory of the tones do not actually overlap (i.e., they come from separate pools of resources, Wickens, 1984; such as those associated with the central executive and the phonological stores in Baddeley's, 1986, working memory model). If so, then one way to extend the present research would be to consider the use of a secondary task involving cognitive resource demands that are more likely to overlap with those used to make comparison responses under speed (where perhaps even an "n-back" tone identification task might serve this purpose). Note, as well, that the fact that secondary task accuracy was around 75% suggests that it was not simply the case that the secondary task was too easy (although the fact that accuracy was at this level for this task might also suggest that participants did not actually try to maintain the tones but simply attempted to recollect them whenever prompted to do so; at which point the resource requirements of the primary task would have had no effect because the comparison response had already been made).

On the other hand, in the present study, a performance decrement in secondary task response time was associated with speeded responding to the primary task. Such a result is indeed somewhat promising because it suggests that some sort of enhanced refractory period might have been occurring after responding under speed to the primary task, which then affected participant's ability to perform the secondary task efficiently (perhaps, by increasing the difficulty of switching to the tone task; Monsell, 2003). Unfortunately, this result is also open to a number of other interpretations which makes it hard to determine the exact locus of this effect. For example, if speeded responding in the primary task was simply associated with enhanced muscle tension, a subsequent slowing of the motor component of response time in the secondary task could certainly then be expected. As well, it is possible that because more errors were occurring in the speeded condition of the primary task, larger response times in a subsequent secondary task might simply represent a form of the well-known post-error slowing effect.

References

Baddeley, A. D. (1986). *Working Memory.* New York: Oxford University Press.

Bogacz, R., Wagenmakers, E., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed-accuracy trade-off. *Trends in Neuroscience, 33*, 10-16.

Brown, S. W., & Merchant, S. M. (2007). Processing resources in timing and sequencing tasks. *Perception and Psychophysics, 69*, 439-449.

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice Hall.

Kleinsorge, T. (2001) The time course of effort mobilization and strategic adjustments of response criteria. *Psychological Research, 65*, 216-223.

Link, S.W. (1992). *The wave theory of similarity and difference*. Hillsdale, NJ: Erlbaum.

Meyer, D. E., Osman, A. M., Irwin, D. E., & Kounios, J. (1988). The dynamics of cognition and action: Mental processes inferred from speed-accuracy decomposition. *Psychological Review, 95*, 183-237.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*, 134–140.

Petrusic, W. M. (1992). Semantic congruity effects and theories of the comparison process. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 962-986.

Ratcliff, R., & Rouder, J. F. (1998). Modelling response times for two-choice decisions. *Psychological Science, 9*, 347-356.

Shoben, E. J., Čech, C. G., Schwanenflugel, P. J., & Sailor, K. M. (1989). Serial position effects in comparative judgments. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 273-286.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological* Review, 108, 550-592.

Van Veen, V., Krug, M. K., & Carter, C. S. (2008). The neural and computational basis of controlled speed-accuracy trade-off during task performance. *Journal of Cognitive Neuroscience, 20*, 1952-1965.

Vickers, D. (1979). *Decision processes in visual perception*. New York: Academic Press.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & R. Davies (Eds.), *Varieties of Attention*. New York: Academic Press.

OLDER AND DISILLUSIONED: AGE DIFFERENCES IN PERCIEVING ILLUSORY LINE MOTION

Vered Shakuf[1] and Yaniv Mama[2]
[1]*Tel Aviv University* [2]*Ariel University Center*
vrdshakuf@gmail.com

Abstract

Older people perform relatively poorly in cognitive-perceptual tasks. Although this finding is widely reported its source is moot. The impairment can result from (a) cognitive origin, (b) sensory origin or from (c) general sluggishness in old age. Speeded tasks are ill suited to resolve the debate because older people respond slower on virtually all tasks. Consequently, in this study we used a non-speeded task of a perceptual judgment. Groups of young and older adults judged the direction of movement of a stationary line that followed the presentation of a dot in an adjacent location (the Illusory Line Motion). The illusion was present with young adults but was absent with older people. We conclude that there are qualitative differences in cognition as a function of age beyond speed of responding.

The sharp increase in life expectancy in the Western world during the last two decades makes the understanding of cognitive-perceptual changes in older age to one of the most important tasks of cognitive sciences. A substantial literature indicates that younger adults generally perform better than older adults on a wide range of cognitive tasks (Poon, 1989).

Age-related decline on the verbal and visuospatial subtests of the Wechsler Adult Intelligence Scale – Revised (WAIS-R) appears to begin as early as age 40 (Lawrence, Myerson & Hale, 1998). Systematic increase of Stroop Interference (SI) with age for adults is often interpreted as reflecting an age-related reduction in selective attention (e.g., McDowd & Shaw, 2000). Research on the effects of aging on adults' ability to interpret spoken language in different situations has demonstrated that older listeners tend to experience greater difficulty understanding speech compared with younger listeners, particularly in challenging environments (Ben David et. al., 2011; also for review, see Schneider, Pichora-Fuller, & Daneman, 2009). Additionally, there are numerous well documented evidences of aging as related to decline in performance on learning and memory tasks (Poon, 1989).

Although these findings are widely reported their source is moot. Age-associated changes in performance are mainly attributed to deterioration in cognitive processes. Thus, for instance, age-associated decline may be due to reduced attention (Stankov, 1988), a failure of inhibition (i.e., a decline in the ability to inhibit irrelevant information; Rabbitt, 1964; Hasher and Zachs, 1988), or reduced information processing speed (Diamond et. al., 2000). However, recent studies have suggested a sensory origin account. Ben David & Schneider (2010) concluded that age-related changes in color perception can contribute to the greater SI observed in aging. In a cross-lab and a cross-sectional analysis, conducted by Ben David & Schneider (2009) they linked sensory losses to the increase in Stroop effects with age.

The vast majority of studies investigating age related differences in cognitive abilities used speeded tasks. These tasks are ill suited to resolve the dispute on the origin of age-related differences because older people respond slower on virtually all tasks. Thus the poorer results of older people may be attributed to mere slower reactions and not to cognitive decline and/or sensory degradation. To shed some light on the abovementioned dispute we used a non-speeded task of a perceptual judgment, The Illusory Line Motion (ILM).

In ILM, a cue appears, and shortly thereafter, a stationary line appears (e.g., Downing & Treisman, 1997). The entirety of the line is presented simultaneously, but observers perceive the line to be presented sequentially, such that the line appears to “unfold” or “be drawn” from the near end of the line (i.e., the end closest to the cue) to the far end (i.e., the end most distant from the cue). There is no actual or implied motion in the display, but observers perceive the far end of the line as moving away from the cue as that line is perceived to extend or expand away from the cue. The most common account of ILM involves differences in the initial strength of attention at different points along the line (Hubbard & Ruppel, 2011). A second notion involving attention and ILM is that additional attention is automatically reallocated (shifted) to the far end of the line following ILM (Hamm & Klein, 2002). Another account of ILM involves apparent motion (e.g., Downing & Treisman, 1997). According to this view, impletion processes involved in apparent motion bind successive presentations of the cue and the line into a representation of a single object. All this accounts of ILM suggests cognitive processes as the source the illusion.

Due to the non-speeded nature of the task, differences in perceiving illusory motion can not be attributed to a general sluggishness of the elderly or deterioration of motor skills. Using high contrast and very discriminable stimuli may suggest qualitative differences in cognition as a function of age beyond speed of responding or sensory degradation.

Method

Participants

The younger participants were 20 undergraduates (mean age M=24.8) from the Ariel University Center, who received partial course credit and were naive to the hypothesis. The older participants were 20 volunteers (mean age M=67.85). All participants were reported to be in good health and had no uncorrected vision problems.

Apparatus

The stimuli were displayed upon and the data were collected with a HP Laptop computer equipped with a 15-in. color monitor with a refresh rate of 60 Hz and a resolution of 1,024 × 768 pixels. Average viewing distance was approximately 60 cm.

Stimuli

Experiment 1 presented an ILM display similar to that in Downing and Treisman (1997) in all respects but the size of the stimuli; a schematic of this display is shown in Fig. 1. First, a fixation point appeared horizontally in the center of the display. A cue then appeared in the left or right side of the screen, and after a brief delay, a horizontal line appeared to the right or left of the cue, respectively, and vertically aligned with the cue. After a brief delay, the display cleared, and a scale for rating the perceived direction of (illusory) motion appeared.

The cue was a black square 60 pixels in width and height (approximately 2.49°), the line was a black rectangle 196 pixels in width and 20 pixels in height (approximately 28.62 × 2.49°), and the cue and line were presented on a white background; the luminance of the cue and of the line was 1.9 cd/m2, and the luminance of the background was 103.0 cd/m2. To rule out sensory differences between the two age groups, both the cue and the line were approximately 9 times (in surface) bigger then the ones used by Downing and Treisman. There was a separation of 60 pixels (2.49°) of empty space between the closest vertical edges

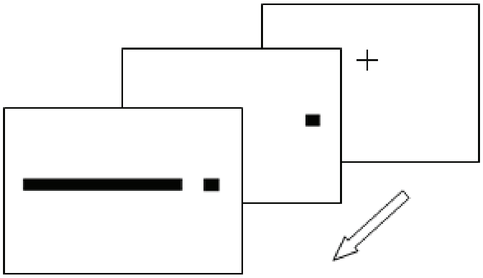


Figure 1: The structure of a trial in Experiment 1. A fixation point is presented (top row), then a cue appears on either the left or right side (second row). A line is then presented (third row). The cue and line vanish (fourth row), and a rating scale appears on the display; the observers rate either velocity or direction (fifth/bottom row).

of the cue and of the line. The SOA between appearance of the cue and appearance of the line was 400 ms. The fixation point was at the center of the display. Ratings of perceived direction were made using a 1–5 scale (in which 1 - was expanded from the right, 2 - was expanded from the left, 3 - was expanded from the middle out, 4 - was expanded into the center, and 5 - was appeared all at once).

Procedure

When participants were ready for a trial to begin, they pressed a designated key. The cue immediately appeared, and the line appeared 400 ms later. The participants responded then by choosing one of the 5 options on the scale. After the participant entered a rating, the display cleared and a prompt to begin the next trial appeared.

Data Analysis

After collecting the data, we transformed the scores using the following index: If participant's response matched cue location (e.g., perceived movement from left to right when the cue was located to the left) the answer was rated +1. If participant's response was opposite to cue location (e.g., perceived movement from left to right when cue was located to the right) the answer was rated -1. Otherwise, (i.e., options 3-5) the answer was rated 0. The meaning of the transformed ratings is that the higher the average score is, the stronger the illusion. Note that guessing all (or most of) the answers will result in an average score close to zero.

Results and discussion

Transformed ratings of perceived direction are presented in Figure 2. The average transformed scores for younger adults was 0.65. This clearly shows that the ILM paradigm induced the illusion of movement for young adults. Whereas, for older adults the illusion disappeared completely and the average transformed scores was -0.01. This difference is significant ($t(19)=12.189$, $p<0.00001$).

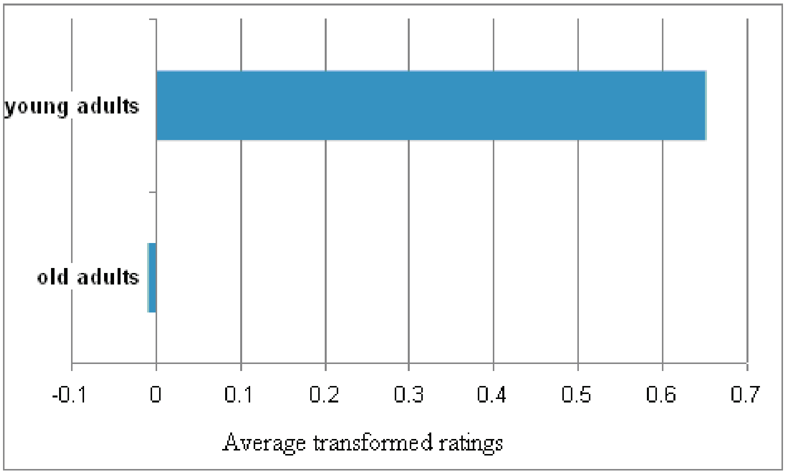


Figure 2: Average of transformed scores. If participant's response matched cue location (e.g., perceived movement from left to right when the cue was located to the left) the answer was rated +1. If participant's response was opposite to cue location (e.g., perceived movement from left to right when cue was located to the right) the answer was rated -1. Otherwise, (i.e., options 3-5) the answer was rated 0.

Since the task did not require any speeded reactions or motor skills the results can not be attributed to slower reactions of older people in general. Furthermore, since we used very large and highly discriminable stimuli, the differences between age groups can hardly be explained by sensory degradation. We conclude that age related differences in ILM can only be attributed to cognitive changes with age.

References

Ben-David, B. M., & Schneider, B. A. (2009). A sensory origin for aging effects in the color–word Stroop task: An analysis of studies. *Aging, Neuropsychology, and Cognition*, 16, 505–534.

Ben-David, B. M., & Schneider, B. A. (2010). A sensory origin for aging effects in the color–word Stroop task: Sim- ulating age-related changes incolor-visionmimics age-related changes in Stroop. *Aging, Neuropsychology, and Cognition*, 17, 730–746.

Ben-David, B. M., Chambers, C. G., Daneman, M., Pichora-Fuller, M. K., Reingold, E. M., & Schneider, B. A. (2011) Effects of aging and noise on real time spoken word recognition: evidence from eye movements. *Journal of Speech, Language, and Hearing Research.* 54, 243-262.

Diamond, B.J., DeLuca, J., Rosenthal, D., Vlar, R., Davis, K., Lucas, G., & Richards, J.R. (2000). Information processing in older versus younger adults: Accuracy vs. speed. *International Journal of Rehabilitation and Health*, *5*, 55– 64.

Downing, P. E., & Treisman, A. M. (1997). The line-motion illusion: Attention or impletion? *Journal of Experimental Psychology: Human Perception and Performance*, 23, 768–779.

Hamm, J. P., & Klein, R. M. (2002). Does attention follow the motion in the "shooting line" illusion? *Perception & Psychophysics*, 64, 279–291.

Hasher, L., and Zacks, R. T. (1988). Working memory, comprehension and aging: A review and new view. In Bower, G. H. (ed.), *The Psychology of Learning and Motivation*, Vol. 22, Academic Press, San Diego, pp. 193–225.

Hubbard, T. L., & Ruppel, S. E. (2011) Effects of temporal and spatial separation on velocity and strength of illusory line motion Attention, *Perception and Psychophys* 73:1133–1146

Lawrence, B., Myerson, J., & Hale, S. (1998): Differential Decline of Verbal and Visuospatial Processing Speed Across the Adult Life Span, *Aging, Neuropsychology, and Cognition*, 5:2, 129-146

McDowd, J. M., & Shaw, R. J. (2000). Attention and aging: A functional perspective. In F. I. M. Craik, & T.A. Salthouse (Eds), *The handbook of aging and cognition* 2nd ed., pp. 221–292. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Poon, L.W. (1989). What do we know about the aging of cognitive abilities in everyday life? In Poon, L.W., Rubin, D. C., and Wilson, B. A. (eds.), *Everyday Cognition in Adulthood and Late Life*, Cambridge University Press, New York, pp. 129–132.

Rabbitt, P. (1964). Ignoring irrelevant information. *British Journal of Psychology* 55(4): 403–414.

Schneider, B. A., Pichora-Fuller, M. K., & Daneman, M. (2009). The effects of senescent changes in audition and cognition on spoken language comprehension. In S. Gordon-Salant, R. D. Frisina, A. N. Popper, & R. R. Fay (Eds.), *The aging auditory system* pp. 167–210. New York, NY: Springer.

Stankov, L. (1988). Aging, attention and intelligence. *Psychology Aging* 3: 59–74.

YOUR NEIGHBORS DEFINE YOUR VALUE: SPATIAL BIAS IN NUMBER COMPARISON

Samuel Shaki [1] and Martin H. Fischer [2]
1. Ariel University Center, Israel
2. University of Potsdam, Germany
E-mail: samuel_shaki@hotmail.com

ABSTRACT

Several chronometric biases in numerical cognition have informed our understanding of a mental number line (MNL). Complementing this approach, we investigated spatial performance in a magnitude comparison task. Participants located the larger or smaller number of a pair on a horizontal line representing the interval from 0 to 10. Experiments 1 and 2 used only number pairs one unit apart and found that digits were localized farther to the right with “select larger” instructions than with “select smaller” instructions. However, when numerical distance was varied (Experiment 3), digits were localized away from numerically near neighbors and towards numerically far neighbors. This neighborhood effect reveals context-specific distortions in number representation not previously noticed with chronometric measures.

When the individual first learns to identify single digits, two-digit numbers are treated in an analytical fashion, so the numbers 19 or 91 have the same values of its two components. Later on, the place value of multi-digits is learned, and the specific location of each digit has its own value and even name (nine or ninety). With increasing experience, people are encouraged to integrate the separate digits into holistic magnitudes. Are single digits of a two-digit number indeed perceptually blended into an inseparable internal magnitude code? Or alternatively, perhaps people are always aware of the constituent values of two-digit numbers.

Three competing theoretical models are suggested to account for the nature of two-digit representations: a holistic, a decomposed, and a hybrid model (Dehaene, Dupoux & Mehler, 1990; Moeller, Huber, Nuerk & Willmes, 2010; Verguts & De Moor, 2005). In its basic form, the holistic model posits that numbers are processed holistically and that people encode two-digit number as an integrated representation before (or along with) executing any comparison task. The decomposed model suggests that the magnitude representations of decade and unit digit are activated separately, either sequentially or in parallel. Finally, the hybrid model suggests that both decomposed as well as holistic representations of number magnitude are activated.

Surprisingly, the multi-dimensional aspect of two-digit numbers has almost never been tested with the traditional methodologies of selective attention. A single study by Fitusi and Algom (2006) explicitly required participants to respond selectively to one of the digits while ignoring the other one. In a Garner paradigm (Garner, 1974), they found that participants could focus their attention on the decade digit only, without suffering interference from an incompatible unit digit. However, participants could not ignore the decade digit when explicitly being asked to do so, demonstrating an asymmetry between both digits of two-digit numbers. Similar asymmetry was found by Ganor-Stern et al., 2007, who argued, “that the syntactic roles of the digits were represented” (p. 483).

The present study has two aims. First, we explored what are the conditions required to turn two adjacent digits into a single two-digit number. Is the mere presence of two digits next to each other sufficient for participants to treat them as a two-digit number? Alternatively, is defining two digits as a two-digit number important? In order to test the above question, we asked participants in one condition to decide the magnitude or parity of two-digit numbers as a whole. In the two complementary conditions, participants were explicitly instructed that the target is the decade or the unit digit only. Looking at the distance effect in these conditions makes it possible to approach the question of two-digit representation from the holistic aspect.

Second, the experimental design permits an examination of the perception of two-digit numbers from the composite side as well. Do people always process the two components of two-digit numbers? Moreover, we are interested in finding whether the decade-unit asymmetry is due to the syntactic roles of the digits (Fitusi & Algom, 2006; Ganor-Stern et al., 2007; Ratinckx, Brysbaert & Fias, 2005) or to the task requirements. Note: parity tasks naturally demand focusing on the unit digit only, while it makes sense to look at the decade digit (first) in comparison tasks. Hence, comparing the compatibility effect between tasks, as well as the decade and unit distance effects, is a test case of the origin of the decade-unit asymmetry.

Method

Participants: Sixteen students (two left-handed males, mean age 22.1 years) from Ariel University Center participated in three experimental sessions, for course credit.

Stimuli and apparatus: The stimulus set consisted of all two-digit numbers in the range 11- 99, excluding numbers with 5 or 0 in it (e.g. 40 or 51). These sixty-four two-digit numbers were replicated twice, resulting in 128 trials in a single block. Half of the stimuli were smaller than 55 and half were larger than 55. Half of the numbers were odd and the other half were even. The stimuli appeared in black Times New Roman font (size 30) at the center of the screen on a white background of a 17-in. (43 cm, 1024 X 768 pixel resolution) monitor. Responses were made on a standard QWERTY keyboard, with all keys covered except A (left hand responses) and L (right hand responses). Reaction times and accuracy were recorded with SuperLab 2.0.

Design: There were two task-order conditions (magnitude task first, parity task first), three target conditions in each task (holistic, decade, unit) and two response rules in each condition (Even-left or Even-right, and small-left or small-right) for each condition, resulting in twelve experimental blocks. This factorial combination (128 trials in each block X 12 blocks) resulted in 1536 trials in total per participant.

Participants performed the holistic conditions of both tasks in the first session. Then, they performed the decade conditions and the unit conditions separately in the two subsequent sessions. The two response rules for each experimental condition were always done in two subsequent blocks. The task order, the decade and unit conditions order, and the response rules were counterbalanced across participants but were kept fixed along the experiment for each participant. Order of trials was randomized in each block.

Procedure: Data were collected in a dimly-lit room with the participant seated approximately 50 cm from the center of the screen. In each trial, a fixation cross appeared for 500 ms on a blank screen and was then followed by a randomly selected two-digit number. In the holistic condition, the participant's task was to decide if the number is smaller or larger than fifty-five (magnitude task), or if the number is odd or even (parity task). In the decade and the unit conditions, the participant's task was to decide if the relevant digit (decade or unit) is smaller or larger than five (magnitude task), or if the relevant digit is odd or even (parity task). The left/right response keys served for odd/even and smaller/larger responses, according to the previously agreed response rule. The stimulus stayed on the screen until a response was made, and the next trial began 500 ms later.

A planned break of 15 minutes between tasks (in the first meeting) and target conditions (in the second and third meetings) was given, while a short 5 minute-break separated between the two subsequent blocks of response rules. Each session took approximately 50 minutes to complete, and approximately one week separated between sessions.

Analysis: RT analyses are presented first, followed by error analyses. Correct RTs were trimmed by accepting responses within ±3 standard deviations, leaving 98.8% of the correct responses for statistical evaluation.

Definition of congruency was determined in a conceptually similar way for magnitude and parity tasks: a two-digit number was defined as congruent if its decade and unit digits lead to an identical response. The number was defined as incongruent if its digits yield conflicting responses. For instance, the two-digit number 34 was considered as congruent in the magnitude task but as an incongruent in the parity task.

Results

Figure1 provides a summary of an ANOVA on RTs, with task (magnitude, parity), the three target conditions (holistic, decade, unit), and congruency (congruent and incongruent numbers). On average, the magnitude task was 29.2ms faster than the parity task [$F(1,15) = 14.44$, $MSe = 2830.3$, $p < 0.01$]. The main effect of target condition was also found reliable: participants performed the holistic condition (644.9 ms) much slower than the decade (609.1 ms) and unit (606.8 ms) conditions [$F(2,30) = 6.01$, $MSe = 4870.5$, $p < 0.01$]. In addition, responding to incongruent numbers was slower (628.7 ms) than to congruent numbers (611.9 ms) [$F(1,15) = 16.56$, $MSe = 815.5$, $p < 0.01$], demonstrating a general failure of selective attention.

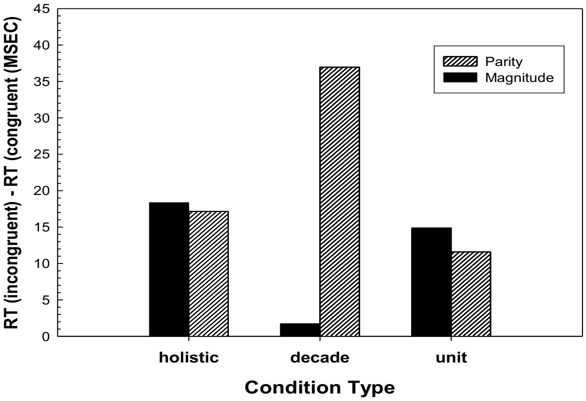


Figure 1. Mean response times (RTs) of congruent numbers subtracted from mean RTs of incongruent numbers (compatibility effect) as a function of condition for comparison and parity tasks.

Did selective attention to the relevant dimension differ between tasks and target conditions? This question was answered by the significant three-way interaction of task, target condition, and congruency [$F(2,30) = 6.06$, $MSe = 309.9$, $p < 0.01$]. As shown in the left panel of figure 1, a significant failure of selective attention was demonstrated in the holistic condition for both magnitude ($t(15)= 2.26$, $p<0.05$) and parity ($t(15)= 1.77$, $p=0.05$). However, when they were asked to focus on the decade digits and to ignore the unit digits (middle panel), participants could turn their attention to the decade digit in the magnitude task ($t(15)= 0.214$, $p>0.05$), but they suffered from intrusions from irrelevant unit digits in the parity task ($t(15)= 5.25$, $p<0.01$). By contrast, interference from the irrelevant decade digit (right panel) was demonstrated in the magnitude task ($t(15)= 2.79$, $p<0.01$), but absent in the parity task ($t(15)= 1.62$, $p>0.05$). None of the two-way interactions turned out to be significant.

The failure of selective attention in some conditions reflects composite processes in those cases. In order to explore possible latter holistic processes (e.g. the hybrid model), we submitted the absolute values of the decade, unit, and global distances to a stepwise multiple regression, in each condition of the comparison task. Regression analyses were highly predictive in all three conditions ($r=0.895$, $r=0.831$, and $r=0.66$, in the holistic, decade, and unit conditions, respectively. All $p<0.001$). Most important, however, is the fact that global distance was the only significant predictor in the holistic condition ($b=-0.895$, $p<0.001$), the decade distance was the only significant predictor in the decade condition ($b=-0.831$, $p<0.001$), and the unit distance was the only significant predictor in the unit condition ($b=-0.66$, $p<0.001$). This pattern of results remained when the logarithms of decade, unit, and global distances were used (e.g. Dehaene et al., 1990; Nuerk et al., 2001).

Finally, the error data generally mirrored the above effects, thereby indicating an absence of a speed–accuracy effect. Overall, error rates were 4.09%, with significantly more errors in the parity task (5.61%) than in the comparison task (2.59%) [$F(1,15) = 5.91$, $MSe = 73.99$, $p < 0.05$]. The main effect of congruency was significant as well, with more errors for the incongruent numbers (4.93%) than for congruent numbers (3.26%) [$F(1,15) = 18.87$, $MSe = 7.1$, $p < 0.001$]. Finally, the three-way interaction of task, condition, and congruent was highly significant [$F(2,30) = 7.86$, $MSe = 5.54$, $p < 0.001$], mirroring again the pattern obtained for RTs: the highest error-rate was found for incongruent numbers in the unit condition (comparison task) and in the decade condition (parity task).

Discussion

Following previous findings of the holistic and decomposed processing of two-digit numbers (e.g. Ganor-Stern, Pinhas & Tzelgov, 2009; Nuerk et al., 2001; Zhang & Wang, 2005), the present study investigated the role of task and condition requirement in such processes.

This study has two important results. First, when introducing two adjacent digits as a two-digit number (e.g. the holistic condition), we found global distance effect and compatibility effect, similarly to previous findings (e.g. Dehaene et al., 1990; 1993; Nuerk et al., 2001). Such results probably reflect holistic and decomposed processes, as was suggested by the hybrid model. By testing participants in the decade and unit conditions, we demonstrated, for the first time, that holistic and even decomposed processes are not obligatory. As the absent of global distance and compatibility effects imply, the mere presence of two digits together is not sufficient to generate holistic or decomposed processes. Instead, participants should be instructed to consider the two digits as a two-digit number.

Second, the asymmetrical pattern of the participant's selective attention indicates the deep involvement of the top-down processes. In the parity task, participants could ignore an irrelevant decade digit but failed to ignore an irrelevant unit digit, probably because in a typical parity task, people have to pay attention only to the unit digit. The opposite pattern was found in the comparison task, most likely because this task usually demands focusing (mostly) on the decade digit.

This later result is in agreement with previous studies (Fitusi & Algom, 2006; Ganor-Stern et al., 2007), which found decade over unit advantage in comparison tasks. However, they concluded that "this pattern of results supports the components with syntactic structure model" (Ganor-Stern et al., 2007, p. 488), in which more weight is given to the decade as compared with the unit digits. The present findings serve as a kind of double dissociation between tasks, and demonstrate how the syntactic structure depends on the natural requirements of the task.

Finally, exploring the distance effects in a comparison task reveals that the global value of two-digit numbers was a significant predictor in the holistic condition only. As is expected from the absence of compatibility effect in the decade condition, the unit and the global distance effects were not significant predictors. Interestingly, the unit distance effect was the only significant predictor in the unit condition, although the failure of selective attention in this condition implies that participants processed the decade digits. Still, the distance effect of decade digits did not emerge, apparently demonstrating dissociation between compatibility and distance effects. It is possible that an irrelevant decade digit is categorically processed (small or large), generating the compatibility effect. However, the exact value of the digit is not projected on the mental number line, resulting in an absence of distance effect.

References

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 624-641.

Fitousi, D., & Algom, D. (2006). Two-digit numbers: How strong is the glue binding their digits? In Kornbrot, D. E., Msetfi, R. M. & MacRae, A. W. (Eds.). Fechner Day 2006. *Proceedings of the Twenty-second annual meeting of the International Society for Psychophysics,* 119-124.

Ganor-Stern, D., Tzelgov, J., & Ellenbogen, R. (2007). Automaticity and two-digit numbers. *Journal of Experimental Psychology: Human Perception and Performance,* 33, 483–496.

Ganor-Stern, D., Pinhas, M., & Tzelgov, J. (2009). Comparing two-digit numbers: The importance of being presented together. *The Quartely Journal of Experimental Psychology*, 62, 444-452.

Moeller, K., Huber, S., Nuerk, H.-C. & Willmes. (2011). Two-digit number processing: Holistic, decomposed or hybrid? A computational modeling approach, *Psychological Research,* 75(4), 290-306.

Nuerk, H. C., Weger, U., & Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. *Cognition, 82*(1), B25-B33.

Ratinckx, E., Brysbaert, M., & Fias, W. (2005). Naming two-digit Arabic numerals: Evidence from masked priming studies. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1150–1163.

Verguts, T., & De moor, W. (2005). Two-digit Comparison: Decomposed, Holistic, or Hybrid? *Experimental Psychology*, *52*(3)*,* 195-200.

Zhang, J., & Wang, H. (2005). The effect of external representations on numeric tasks. *The Quarterly Journal of Experimental Psychology, 58A*(5), 817–838.

THE SPATIAL REPRESENTATION OF NON-SYMBOLIC NUMERICAL QUANTITIES

Riccardo Luccio, Antonia Fumarola, Giorgia Tamburini, & Tiziano Agostini
Department of Life Sciences, University of Trieste

Abstract.

Dehaene et al. (1993) have demonstrated an association between number magnitude and response position (SNARC effect). More recently, this association has been demonstrated also for non-symbolic quantities (see Shaki, Petrusic & Leth-Steensen, 2012). de Hevia & Spelke (2009) have used different numbers of dots as stimuli in a bisection task, showing a SNARC-like effect (but see Gebusi & Gevers, 2011). We investigated the association between representation of non-symbolic numerical quantities (dots) and response position in a simple detection experiment (see Fischer, 2003). The dots were used as prime, and the participants had to press a bar, as soon as they detected a grey square on the right or on the left. Our results (faster RTs for small quantities and target on the left, and for big quantities and target on the right) support the idea of a mental representation of non-symbolic quantities as a left-to-right oriented mental line.

Dehaene, Bossini and Giraux (1993) reported a spatial relation between number magnitude and its representation (i.e., Spatial Numerical Association of Response Codes: SNARC effect). In identification or comparison of numbers, participants were faster at processing large numbers (e.g., 9, presented in the center of a screen), when responses were executed in the right hemispace, whereas they were faster at judging smaller numbers (e.g., 1 presented in the center of a screen), when responses were executed in the left hemispace. The SNARC effect suggests that the representations of relatively small numbers magnitudes are spatially compatible with the left hemispace and those of relatively large numbers magnitudes are spatially compatible with the right hemispace (i.e., left-to-right oriented mental number line: MNL). This effect has been shown not only for number magnitude, but also for non-numerical ordered sequences. Gevers, Reynvoet, and Fias (2003) investigated the spatial organization of two non-numerical ordered sequences: names of the months, and letters of the alphabet. Gevers et al. asked participants to judge whether months presented in the centre of a screen came before or after "June", and to judge whether letters presented in the centre of a screen came before or after the letter "O". Results showed that the mental representation of these ordinal sequences could be spatially coded, because the names of the first months of the year were processed faster with responses executed in the left hemispace, whereas the reverse pattern was obtained for the last months of the year. Gevers et al. reported similar findings also on the task employing letters. In a similar vein, Rusconi et al. (2006) showed that even pitch is represented along a mental line. They explored the spatial representation of pitch height through the pairing of pitch to different response positions. In the first task, non-musicians were asked to compare the frequency of two pure tones. In the second task, non-musicians and musicians were asked to classify sounds as being produced by wind or percussion instruments. Results showed that the internal representation of pitch height was spatially organized, especially in participants with formal musical education (i.e., Spatial Musical Association of Response Codes: the SMARC effect). For a review of the SNARC effect, see Hubbard, Piazza, Pinel, and Dehaene (2005); for recent results on non-numerical magnitudes, and an up-to-date review, see Shaki, Petrusic, and Leth-Steensen (2012).

According to Walsh (2003), who has proposed a unifying framework called the ATOM (A Theory of Magnitude), the SNARC effect might be better understood as an instance of the SQUARC effect (Spatial Quantity Association of Response Codes), whereby any magnitude that is coded spatially or in action implies a relationship between magnitude and space. It follows that experiments in which responses are made to two or more magnitudes (independently of how they are coded), should show a magnitude priming on successive trials, regardless of the domain of coding. In this vein, de Hevia and Spelke (2009) have demonstrated, with a bisection task, that children show a SNARC-like effect with non symbolic (dots) representations of quantities, as well as for symbolic representations. However, Gebuis and Gevers (2011) have cast some doubt on those results, demonstrating that de Hevia and Spelke neglected a possibly confounding variable, that is, the extent of the area occupied by the dots (larger when the dots are more numerous).

On the basis of the studies by de Hevia and Spelke, and of Walsh's ATOM, we hypothesize that there is a spatial representation not only of the numbers, but also of the amount that they represent. We know from the literature that non-symbolic quantities are processed differently, depending on the number of items present. Amounts ranging up to 6-7 elements can be calculated very quickly, at a glance. This type of processing is called *subitizing*. On the other hand, amounts ranging from 9 elements onwards require a more accurate calculation; they cannot be calculated by means of a glance but must be counted. This process is called *counting*. In our study, we used the quantities that fall within the range of subitizing and of the counting to see how non symbolic quantities are spatially represented both in the range of subitizing and of counting.

EXPERIMENT

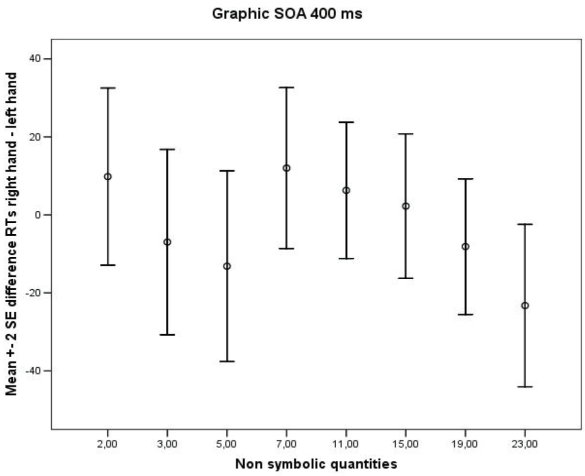
Method

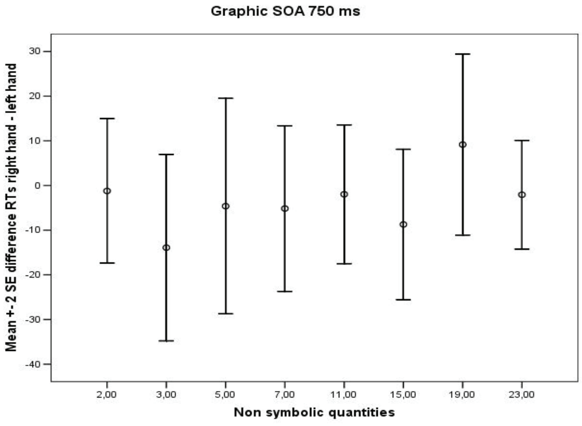
Sixteen students participated in the experiment, 1 male (age: 31 years) and 15 females (mean age: 21.5 years). Fourteen participants were right-handed, one was left-handed and only one was ambidextrous. All subjects had normal or corrected-to-normal vision. We used E-Prime software (Version 1.2., http://www.pstnet.com/eprime.cfm) to create and administer the experiment. Stimuli were displayed on a 26-inch (Quato Intelli Proof), with a 1024 x 768 pixel resolution (screen refresh rate: 75 Hz). The PC was a Dell Optiplex 382 Intel core (2) (RAM: 512 Mb), running Windows XP. We used a response box to record participants' responses. Each trial was comprised of three stimuli. The first stimulus was a fixation cross measuring 1° by 1°, that was placed between two boxes (dark grey). The second stimulus was an array of black dots (2, 3, 5, 7, 11, 15, 19, 23) that were presented in the middle of a grey circle. The third stimulus was a square target (light grey) that appeared in one of the two boxes. All stimuli were presented on a black background.

The experiment took place in a quiet, dimly lit room without environmental distractions. Participants sat in front of the monitor and were asked to put their index finger on the centre key of the response box. The viewing distance was 57 cm. Each trial started with a white fixation cross displayed for 300 ms at the centre of the screen, between two grey boxes, followed by an inter-stimulus interval (ISI) of 130 ms, consisting of a black screen. Then, the dots appeared for 200 ms. After an SOA, one of the two light grey target squares appeared for 1000 ms. Participants were asked to try to estimate the number of dots in their mind and then to press the centre key on the response box, when the target appeared. The inter-trial interval (ITI) was 1500 ms. The experiment was comprised of two sessions. In the first session, participants were asked to press the key with their right index finger when the square target appeared. In the second session, participants responded with the opposite hand. There was a short break between sessions, and the order of sessions was counterbalanced across participants. Each session comprised three blocks of trials (i.e., the training block and two experimental blocks). Each session started with the training block of eight trials. Thereafter, in each experimental block, the eight quantities were presented for ten times in random order (for a total of 160 trials).

Results

The data were analysed with regression for repeated measures (Fias et al., 1996; Lorch & Myers, 1990). The independent variable was the quantity values and the dependent variable was the difference between the median reaction time (RT) of the left target and the median RT of the right target: dRT = RT(right target) - RT(left target). In the first step, for each participant the median RT of the responses was computed for each quantity level, separately for left and right hand responses. Then, dRT was computed by subtracting the median RT of left target responses from the median RT of right target responses. In the second step, a regression equation was computed for each participant. In the third step, one-sample t-test were performed to test whether regression beta weights of the group deviated significantly from zero. The analysis revealed that the regression slopes (regression beta coefficients) were significantly different from zero, $t(15) = -2.697$, $p < .05$, only for the responses to the 400 ms SOA and for quantities within the range of counting. There was a left target advantage in processing the quantities 7-11 and a right target advantage in processing the quantities 19-23. These results can be interpreted as evidence that the non symbolic quantities in the range of counting were spatially represented on a mental line oriented from left to right (see the figure – top, SOA 400 msec, bottom SOA 750 msec).


Graphic SOA 400 ms


Graphic SOA 750 ms

Among other analyses, we have also performed an ANOVA for SOA at 400 ms and SOA at 700 ms, both for the quantities in the range of subitizing, and for the quantities in the range of counting. The results confirm those of the regression analyses, showing a significant interaction between target position and quantity, in the range of counting ($F(1, 15) = 7.272, p < .05$).

CONCLUSION

The SNARC effect shows that the numbers are mentally represented over the space, along a line oriented from left to the right (Dehaene et al., 1993). In the last few years, it has been demonstrated that this kind of representation holds not only for numbers, but also for other concepts (Rusconi et al. 2006, Shaki et al. 2012). Recently, it was asked if the same kind of spatial representation applies also to non-symbolic quantities, using a line bisection task. The evidence, however, is far from being unambiguous. The aim of our study was to determine whether the quantities given by dot patterns can be represented spatially in the range respectively of *subitizing* and *counting*. What we have found is that only in the range of counting are these quantities represented spatially according to a mental line that goes from left to right. Moreover, this effect is present only with a SOA of 400 ms, and not with a SOA of 750 ms. Probably this could be due to the fact that the SNARC effect, as shown by Fischer (2003), is one that becomes manifest after times of about 400-500 ms SOA. These results agree with the ones obtained by de Hevia and Spelke (2009), but do not answer to the objections raised by Gebuis and Gevers (2011), which require more research.

REFERENCES

de Hevia, M. D. & Spelke (2009). Spontaneous mapping of number and space in adults and young children. *Cognition, 110(2)*, 198–207.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity & number magnitude. *Journal of Experimental Psychology: General, 122*, 371–396.

Fias, W., Brysbaert, M., Geypens, F., & d'Ydewalle, G. (1996). The importance of magnitude information in numerical processing: Evidence from the SNARC effect. *Mathematical Cognition, 2,* 95-110.

Fischer, M.H. (2003) Spatial representations in number processing – evidence from a pointing task. *Visual Cognition, 10,* 493–508.

Gebuis, T. & Gevers (2011).Numerosities and space; indeed a cognitive illusion!, *Cognition, 121*. 248–252

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition, 87*, B87–B95.

Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews and Neuroscience, 6*(6), 435–448.

Lorch, R., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 149-157.

Rusconi, E., Kwan, B., Giordano, B., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: The SMARC effect. *Cognition, 99, !*13-129.

Shaki, S., Petrusic, W. M., & Leeth-Steensen, C. (2012). SNARC Effects With Numerical and Non-Numerical Symbolic Comparative Judgments: Instructional And Cultural dependencies. Journal of Experimental Psychology:*: Human Perception and Performance, 38*, 525-530.

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, 7, 483–488.

EXTENDING SNARC: FROM SINGLE NUMBERS TO SEQUENCES

Dean A. Verger and William M. Petrusic
Carleton University, Ottawa, Ontario, Canada
dverger@connect.carleton.ca
bill_petrusic@carleton.ca

Abstract

We found a Spatial Numerical Association of Response Codes effect for sequences of three single-digit numbers presented in English when Chinese speaking students determined if the sequence was in an ascending order.

People develop their understanding of numbers over time and experience. They begin with approximations and then learn to associate a single label with a specific quantity. A pre-school aged child playing with three erasers may know the word three corresponds to the three erasers, but may not yet understand that two comes before three, and four after. They eventually learn that numbers fit into a number line, both through recitation, and the visual reminder, often seen in elementary school class rooms. Siegler and Opfer (2003) showed that number ordering strategies change with age developing from either a log or a power function in grade 2 to a linear function after grade 6 through adulthood.

Dehaene, Bossini, and Giraux (1993), in a series of experiments, provided evidence for a mental representation of a number line. They found a magnitude of number by side of response interaction when deciding parity, this in spite of the fact that the magnitude of the number was irrelevant to the task. The time to respond was faster for small numbers when the side of response was on the left rather than the right. Conversely, the time to respond was faster for large numbers when the side of response was on the right rather than on the left. Coined the Spatial Numerical Association of Response Codes (SNARC) effect, it is typically reported as a difference in response times (dRT) between the left sided and right sided responses. They initially reported that magnitude, and not order, was the source of this new effect, as an experiment using letters as stimuli failed to elicit a SNARC effect.

The SNARC effect has been found to be affected by the direction of reading, the range of the number set used, and contrary to the original findings, has been extended to non-magnitude stimuli. Gevers, Reynvoet, and Fias (2003) found a SNARC effect for months of the year and letters of the alphabet (within order relevant and irrelevant tasks) and the days of the week (Gevers, Reynvoet, & Fias, 2004).

Previtali, de Hevia, and Girelli (2010) showed that a learned order of words could also elicit a SNARC-like effect. They interpreted this as the newly learned ordered sequences being learned in the same way as numbers and letters, and as such were represented in an ordered manner in spatial memory.

Prado, Van der Henst, and Noveck (2008) used relational reasoning to indirectly develop a sequential mental array of a seating plan. They found that the decisions about intermediate positions were noisy, that there was too much interference between the stimuli in the middle of the sequence. They therefore used only the end points in a subsequent analysis, and found a significant SNARC effect.

But not everyone accepts the mental number line as the under lying reason for the SNARC effect. Santens and Gevers (2008) argue for a three level model, wherein the intermediate level codes result in the effect.

If it is the case that we create an analog mental representation of learned sequences that maintains a culturally appropriate order, does this occur with groupings of numbers? In the following experiment sequences made up of three non-repeating numbers were presented. If the SNARC effect arises out of a correspondence between number position on a mental number line, then does it also extend to sequences of numbers when those sequences are made up of only small numbers or only large numbers?

Method

Participants

Thirty-four Carleton University students participated for partial course credit. Three student's data were removed from analysis due to non-performance of the tasks. The results from 19 women and 12 men with a mean age of 21.8 years (2.7 sd) were included in the analysis. Participant first-language breakdown: Arabic (6), Chinese (11), English (14). Three students were left-handed.

Equipment

A Pentium computer, a MultiLink LCD monitor, and a modified KeyTronic keyboard were used. Two keyboard keys were enabled as targets: A on the left side, and L on the right, with the other keys removed from the keyboard to avoid accidental key presses. The stimuli were controlled and the results collected using Superlab version 4.5 software. All instructions and stimuli were presented in Tahoma, regular, 20 point font, centered on the screen.

Stimuli

Stimuli triads were made up of three non-repeating single digit numbers from the number set {1,2,3,4,6,7,8,9}. Triads were categorised as in Table 1 These three numbers were spaced so as to be read as a single number made up of hundreds, decades and units. The practice also emphasised the separate and sequential nature of the numbers. These two sequences thus corresponded to the beginning and end of the single digit number set.. Category 1 composed of ordered and mixed triads from the subset of {1 2 3}. Category 2 were made up of {1 2 3 4} but not including Category 1. Category 3 included two numbers below 5, and one above. Category 4 included one number below 5 and two above. Category 6 was composed of the number set {7 8 9}. Category 5 was made up of numbers above 5 that did not include Category 6. The order of presentation of the tasks (Ascending, Descending) and side of response were counterbalanced between participants. Ascending had no descending sequences. Descending had no ascending sequences

Procedure

A centered fixation *   *   * was displayed for 1000 ms followed by a three-number sequence. If the stimulus triad was an ordered sequence the participant pressed either the right or the left button on the keyboard according to the instructions. The fixation then reappeared in the middle of the screen in preparation for the next trial. Three blocks of each condition were presented with a self-determined break. Six practice trials with feedback were done four times during the experiment, once before each new task. There were 72 stimuli x 3 blocks x 2 sides x 2 tasks = 864 responses. The experiment took approximately 45 minutes. Participants were told to be as fast and as accurate as possible.

Table 1. Examples of the three-number ordered and mixed stimuli. One and six are the end points of the Size factor. Size 1 & 2 contain non-repeating numbers under 5. Size 5 & 6 contain non-repeating numbers over 5. Size 3 & 4 contain non-repeating numbers crossing 5.

Type	1	2	3	4	5	6
Ordered	{1 2 3}	{1 3 4}	{3 4 7}	{3 7 8}	{6 7 9}	{7 8 9}
Mixed	{2 1 3}	{3 4 1}	{4 3 7}	{7 3 8}	{7 6 9}	{8 7 9}

Results

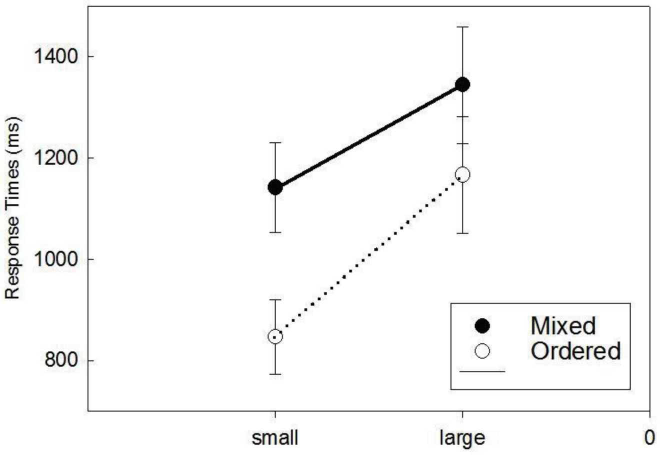


Figure 1. Interaction plot of Size (for the end points) and Ordered (ordered or mixed sequence). Errors bars are standard errors.

Incorrect responses were removed from the analysis (5.3% errors). Response times over the Mean RT plus 3 times the Standard Deviation for each participant were also dropped (1.9%) leaving 92.8% of the responses to be analysed.

A mixed model ANOVA was conducted with Language, Gender, and Handedness as between subjects variables. The main effects of Gender, Handedness, Language, Side of response, and Direction of Task were not significant. The main effect of Size was significant. Small triplets were responded to more quickly than large triplets, $F(1, 30) = 45.17$, $p < .00$. The main effect of Order was significant. Ordered triplets were responded to more quickly than mixed triplets, $F(1, 30) = 102.06$, $p < .001$.

The interaction of Size x Order was significant. As in Figure 1, ordered sequences were responded to more quickly than mixed sequences, and large three-number sequences took more time to respond to than small three-number sequences, $F(1, 30) = 8.34$, $p = .007$. The interaction of Size x Order x Direction trended towards significance, $F(1, 30) = 3.19$, $p = .084$.

SNARC

The Size x Side of Response x Direction (task) interaction was significant, $F(1, 30) = 4.34$, $p = .046$. The difference in response times (dRT) was calculated for each end point, using Right – Left hand response times for each participant (Figure 2). Individual regressions were run and tested for the Size by Side of response dRTs for Ascending, and for Descending following the process recommended by Lorch and Myers (1990). The single sample *t*-test for Ascending instruction slopes was not significant, $t(30) = -0.680$, $p = .501$. Whereas, for Descending it trended towards significance, $t(30) = -1.724$, $p = .095$.

Because the SNARC effect is dependent on the direction of reading and writing, a new set of single-sample *t*-tests was run taking writing direction into account. For participants who wrote left to right in their first language the regression slopes for dRT end points for Ascending instruction were still not significant, $t(24) = -0.650$, $p = .522$, 2-tailed. Descending regression slopes continued to trend towards significance, $t(24) = -1.967$, $p = .061$. For participants whose writing in their first language was right to left, Ascending, $t(5) = -0.222$, $p = .833$, 2-tailed, and Descending, $t(5) = -0.179$, $p = .865$, 2-tailed, regression slopes were not significant.

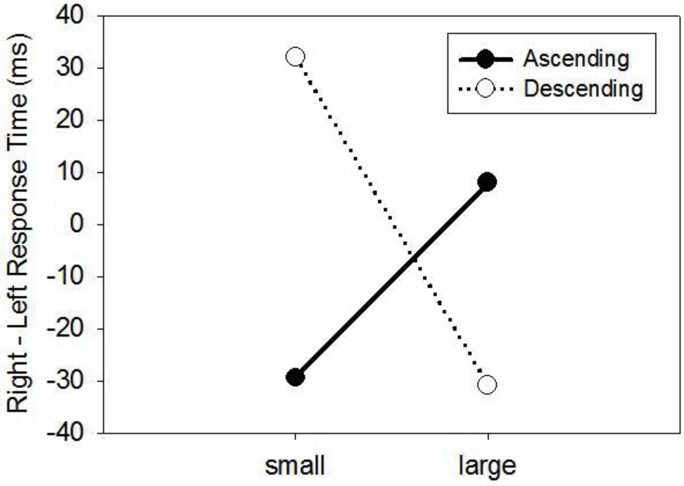


Figure 2. Interaction plot showing the end point dRTs for Ascending and Descending sequence instructions for all participants together

A further breakdown by language was conducted to determine if language groups had different patterns of responses to the stimuli. Table 2 shows the slope directions and significance testing results for each language group. The single-sample *t*-test was only significant for the Chinese language group for the Ascending instruction, *t*(10) = -2.698, *p* = .022.

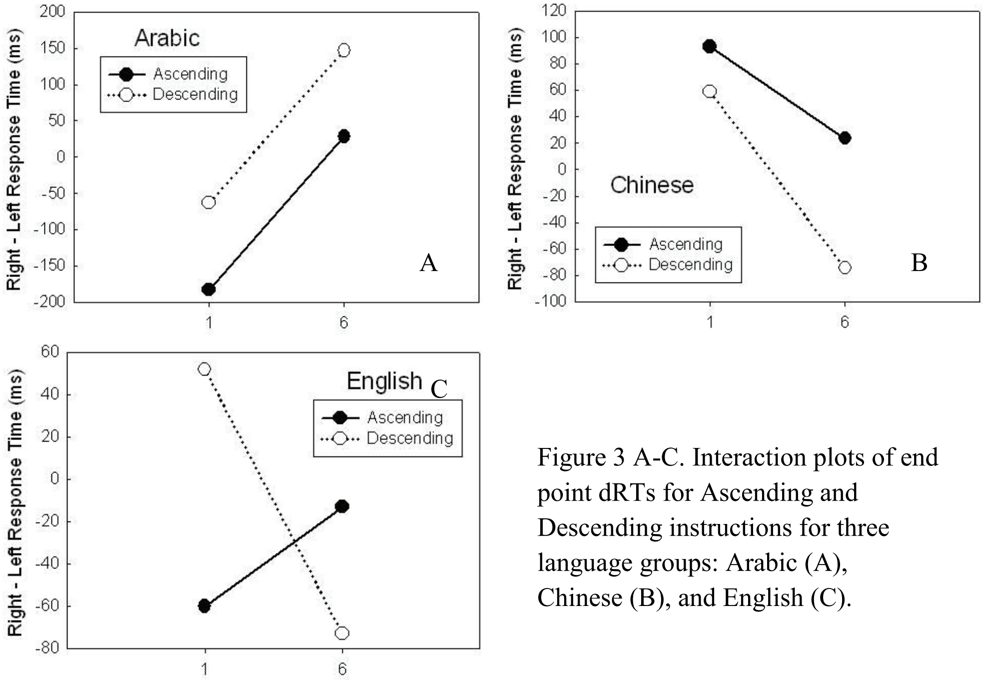


Figure 3 A-C. Interaction plots of end point dRTs for Ascending and Descending instructions for three language groups: Arabic (A), Chinese (B), and English (C).

Table 2. Single-sample *t*-test (2-tailed) of individual participant regression slopes for both instructions by language for each directional task.

Language	*df*	*t* value	*p* value
Arabic			
Ascending	5	-0.222	.833
Descending	5	-0.179	.865
Chinese			
Ascending	10	-2.698	.022
Descending	10	-1.394	.194
English			
Ascending	13	1.675	.118
Descending	13	-1.336	.204

Discussion

The main focus of this experiment was to determine the explicit effect of order on the SNARC effect. Rather than have a single digit number, triplets were used (combinations of three single-digit numbers). A global SNARC effect was found at the end points of the possible sequences (the smallest or the largest numbers of the number set).

In the initial analysis, there appeared to be a crossover effect between the two tasks: deciding if the sequence was Ascending (Descending). Interestingly, the slopes were in the opposite directions to what would have been expected (Figure 2). In a regular SNARC effect, one might have expected that the left hand responses would have been faster for the small number sequence, and the right hand faster for the large number sequence within the Ascending task. Instead it appeared that for the all-participant analysis the Ascending task the dRT slope was positive, rather than negative. This pattern was reversed for the Descending task.

However, in the first two analyses, the full analysis, and the analysis based on reading direction, only the Descending task had a dRT slope that approached significance. This led to the third analysis broken down by first language of participants (Table 2). Only the dRT regression slopes for the Chinese students within the Ascending decision task passed the Lorch and Myers (1990) test.

Of interest are the different dRT slopes for the two tasks within the three languages as seen in Figure 3. For the dRT slopes for the English participants, the language of the experiment, there was a crossover effect. The English participants attended to the differences in the tasks changing their framework from which to make decisions. However, as in the overall analysis the expected slopes of the dRT are reversed, with Ascending being positive, and Descending negative.

One explanation for this apparent reversal might be understood through the process of reading. Reading tracks from left to right. With the reading of the third number a decision is made. The focus now is to the right, rather than the center of the screen. The decision is thus anchored on the right side, essentially flipping the framework. Rather than left to right, the decision is based on a right to left situation.

This explanation can also be applied to the Descending decision task. Since descending reverses the number line (9 8 7 6 …) the same process of reading left to right but deciding right to left flips the reversed number line, resulting in a dRT slope that is negative.

This interpretation works well for Figure 2 and the English component of Figure 3 where the dRT slopes for Ascending is positive and Descending negative. In the case of the Arabic students both Ascending and Descending slopes are positive. Reading direction in this language is right to left. Because only six Arabic students participated it is difficult to interpret what might be happening. The dRT slope for Descending is the reverse of that shown in English, as expected. But the dRT slope for Ascending is the same. It is possible that the students were using a mix of frameworks, with Ascending in English, but the more difficult instruction/task being recoded in Arabic.

The dRT slopes for both tasks for the Chinese students were negative. The dRT slope for Ascending was the reverse of the slope as in the English language analysis. Chinese students use the same left to right reading direction as English students. In this case the argument for task difficulty leading to a flipping of the framework cannot be made. It may be that Chinese students, who have learned their numbers through rote, treat Ascending and Descending in the same manner, without a reversal of the spatial framework.

The dRT slope for the Descending task was in the same direction for Chinese and English participants. This directional compatibility may be the reason why a SNARC effect was first noted for the Descending task.

Finally, only the analysis of slopes for the Ascending task for the Chinese students reached significance. Within this task, the Chinese students were faster at responding to whether a small number sequence was in order with their left hand, and faster with their right hand for a large number sequence.

Conclusion

A global SNARC effect was found for responses by Chinese students who were asked to determine if a sequence of three numbers were in ascending order.

Interestingly, patterns of dRTs varied by language. Unfortunately, this could not be further investigated due to the loss in power when the participant pool was subdivided into smaller language groups. Future research should look at increasing the number of participants within each language. It should also determine if difficulty of task has an impact when the task needs to be translated.

References

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122*(3), 371-396.

Gevers, W., Reynvoet, B., & Fias, W. (2004). The mental representation of ordinal sequences is spatially organised: evidence from days of the week. *Cortex, 40*, 171-172.

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition, 87*(3), 87-95.

Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition. 16(1)*, 149-157

Prado, J., Van der Henst, J., & Noveck, I. A. (2008). Spatial associations in relational reasoning: Evidence for a SNARC-like effect. *The Quarterly Journal of Experimental Psychology, 61*(8), 1143-1150.

Previtali, P., de Hevia, M. D., & Girelli, L. (2010). Placing order in space: The SNARC effect in serial learning. *Experimental Brain Research, 201*(3), 599-605.

Santens, S., & Gevers, W. (2008). The SNARC effect does not imply a mental number line. *Cognition, 108,* 263-270.

Siegler, R.S., & Opfer, J. E. (2003). The development of numerical estimation: evidence for multiple representations of numerical quantity. *Psychological Science, 14(3),* 237-243.