FECHNER DAY 2017

Conference Proceedings

The 33rd Annual Meeting of the International Society for Psychophysics

Sunday to Thursday, 22–26 October 2017

Kyōsō-kan and Hon-kan, Denki Building, Fukuoka, Japan

Organized by

Research Center for Applied Perceptual Science (ReCAPS)
Kyushu University

ReCAPS

Sponsored by

FCVB 公益財団法人 福岡観光コンベンションビューロー
Fukuoka Convention & Visitors Bureau

KYUSHU UNIVERSITY

Proceedings of the 33rd Annual Meeting of the International Society for Psychophysics

Fechner Day 2017

Fukuoka, Japan
22–26 October 2017

How to cite this book: Kazuo Ueda, Hiroyuki Ito, Gerard B. Remijn, and Yoshitaka Nakajima (Eds.) *Fechner Day 2017: Proceedings of the 33rd Annual Meeting of the International Society for Psychophysics.* International Society for Psychophysics, Fukuoka, Japan, 2017.

Distribution: International Society for Psychophysics, c/o Kazuo Ueda, Department of Human Science, Kyushu University, 4-9-1 Shiobaru Minami-ku, Fukuoka 815-8540, Japan.

LaTeX typeset by Kazuo Ueda.

Printed by Kijima Printing Inc., 2-9-6 Shirogane, Chuo-ku, Fukuoka-shi, Fukuoka 810-0012, Japan.

International Society for Psychophysics

Founded in Cassis, France, 22 October 1985.

Executive Committee 2017–2018

Organizing Committee of Fechner Day 2017

Program Committee of Fechner Day 2017

Public Relations Committee of Fechner Day 2017

Advisory Board of Fechner Day 2017

Past Annual Meetings

Year	Venue	Organizer(s)
1985	Marseille, France	G. Canévet and B. Scharf
1986	Cassis, France	A. M. Bonnel, G. Canévet, C.-A. Possamaï, B. Scharf, B. Berglund, U. Berglund, and R. Teghtsoonian
1987	Durham, NH, USA	M. Teghtsoonian and R. Teghtsoonian
1988	Stirling, Scotland	H. Ross, R. Macdonald, C.-A. Possamaï, R. Teghtsoonian, M. Treisman, and R. Warren
1989	Cassis, France	A. M. Bonnel, G. Canévet, C.-A. Possamaï, and B. Scharf
1990	Würzburg, Germany	F. Müller
1991	Durham, NH, USA	G. R. Lockhead
1992	Stockholm, Sweden	G. Borg, Å. Hellström, and G. Neely
1993	Mallorca, Spain	A. Garriga-Trillo, P. R. Miñón, C. Carcía-Gallego, P. Lubin, J. M. Merino, M. J. Rubio-Gómez, and A. Villario
1994	Vancouver, Canada	L. M. Ward
1995	Cassis, France	C.-A. Possamaï, H. Ross, B. Scharf, R. Teghtsoonian, and L. M. Ward
1996	Padua, Italy	S. C. Masin
1997	Poznan, Poland	R. Gotebiewski, E. Hojan, T. Hornowski, P. Kokowski, A. Majchrzak, P. Miecznik, M. Labowski, P. Pekala, A. Preis, E. Skrodzka, E. Wichlinska, and A. Wicher
1998	Québec, Canada	S. Grondin, Y. Lacouture, and R. Rousseau
1999	Tempe, AZ, USA	P. R. Killeen and W. R. Uttal
2000	Strasbourg, France	C. Bonnet
2001	Leipzig, Germany	E. Sommerfeld, R. Kompass, and T. Lachmann
2002	Rio de Janeiro, Brazil	J. A. da Silva, E. H. Matsushima, and N. P. Ribero-Filho
2003	Larnaca, Cyprus	B. Berglund and E. Borg
2004	Coimbra, Portugal	A. Olivera, M. Teixeira, G. F. Borges, and M. J. Ferro
2005	Traverse City (MI), USA	J. S. Monahan, S. M. Sheffert, and J. T. Townsend
2006	St Albans, UK	D. E. Kornbrot, R. M. Msetfi, and A. W. McRae
2007	Tokyo, Japan	S. Mori, T. Miyaoka, and W. Wong
2008	Toronto, Canada	B. A. Schneider, B. M. Ben-David, S. Parker, and W. Wong
2009	Galway, Ireland	M. A. Elliott, S. Antonijević, S. Berthaud, P. Mulcahy, C. Martyn, B. Bargary, and H. Schmidt
2010	Padua, Italy	A. Bastianelli, G. Vidotto, L. Filipponi, D. Polezzi, A. Spoto, D. Massidda, and M. Scandola
2011	Herzliya, Israel	D. Algom, D. Zakay, E. Chajut, S. Shaki, Y. Mama, and V. Shakuf
2012	Ottawa, Canada	S. R. Carroll, E. Gallitto, C. Leth-Steensen, W. M. Petrusic, J. R. Schoenherr, and D. A. Verger
2013	Freiburg, Germany	J. Wackermann, M. Wittmann, and W. Skrandies

Year	Venue	Organizer(s)
2014	Lund, Sweden	G. R. Patching, Å. Hellström, M. Johnson, E. Borg, and R. Bååth
2015	Québec, Canada	S. Grondin, V. Laflamme, G. Plante, L. Azari, and M.-L. Roy
2016	Moscow, Russia	I. G. Skotnikova, A. N. Gusev, G. Y. Menshikova, I. Blinnikova, O. Korolkova, S. A. Emelianova, V. M. Shendyapin, V. E. Doubrovski, A. I. Kovalev, A. Kremlev, and N. N. Volkova

Contents

Welcome to Fukuoka!

This is the second time an annual meeting of the International Society for Psychophysics—a Fechner Day—has taken place in Japan, the first being the Fechner Day in Odaiba, Tokyo, in 2007. However, this is the first on Kyushu Island, of which Fukuoka is the political and economic center. Throughout history, Fukuoka and the surrounding neighborhood has played an important role in establishing international relations as well as educating students and monks.

In the middle of the first century, a diplomatic team of the small kingdom of *Na,* which was located in or near Fukuoka, visited the court of the Later Han Dynasty (25 A.D.–220) in China, and received a golden stamp (Fig. 1) as an imperial present. This golden stamp now belongs to the Fukuoka City Museum. It was lost for a long time, buried in a rice field, and was discovered in the 18th century. A local scholar, called Nanmei Kamei (1743–1814), identified it as the stamp described in a historical text. He was a great teacher, and one of his students, Tanso Hirose (1782–1856; Fig. 2), opened a school in Hita, an attractive city two hours by bus from Fukuoka. His school was the first educational institution in Japan that accepted students from any social class and any gender.

People in Fukuoka are open-minded, and the northern part of Kyushu has been an important area for the cultivation of academic freedom in Japan. Academic freedom was often combined with nice food and drinks.

We would like to make the present Fechner Day an enjoyable occasion. We have booked a large conference room for the first four days, and this will be a place for our close communication. We have to move to the basement of a neighboring building for the last day's proceedings, and I hope that participants will get accustomed to this neighborhood by then. There are many attractive places to eat and drink near the conference site, and prices are in many cases inexpensive. Keep in your mind that you never need a tip in Japan—at least during your stay.

We will try two things to make the present Fechner Day special. The poster symposia, with some snacks and drinks, will be ideal occasions to get acquainted with

Fig. 1. The golden stamp brought to Kyushu in 57 A.D. according to the literature. This stamp was discovered in Fukuoka in 1784. Photos courtesy of the Fukuoka City Museum.

Fig. 2. Tanso-sensei with me.

one another. Please, follow the chairperson's instructions at the beginning. There are keynote lectures and symposia showing how our place is connected to other places of the world. Since our place is psychologically—psychophysically?—far from any of the places in which previous Fechner Days have been organized, this will be a good opportunity for many participants to get an impression about what is happening here in Fukuoka and in other places connected to Fukuoka.

The organizing team members (members of all committees) are grateful to the executive board of the International Society for Psychophysics, previous Fechner Day organizers, and the Advisory Board for their advice and support. Many people worked for this occasion in many different ways, and I thank all of them heartily. Finally, the participants are thanked for making the occasion so exciting.

Please, have fun, and find and think a lot of interesting things in Fukuoka!

Yoshitaka Nakajima, PhD

Chair, Fechner Day 2017
Research Center for Applied Perceptual Science
Faculty of Design
Kyushu University

Part I

Keynote Lecture 1

VISUAL PSYCHOPHYSICS WITH NATURAL IMAGES

Isamu Motoyoshi
Department of Life Sciences, The University of Tokyo, Tokyo, Japan
<motoyosilab@gmail.com>

Humans can easily recognize objects, scenes, and faces in natural environments. The retinal image of 3D objects and scenes is highly complicated, and therefore many psychophysical studies have scrutinized the visual perception elementary features such as shape and color by using simplified stimuli on the assumption that analyzing such features is necessary for visual cognition in general. Psychophysical investigations, in concert with physiological evidence that neural sensors detect such images features, have certainly improved our understanding of the visual system. But can this assumption of lower-level elementary feature analysis be generalized to higher levels of visual cognition? For instance, does the attractiveness of an apple require the apple to be recognized, which in turn may necessitate the reconstruction of the apple's 3D shape and surface reflectance, which may itself imply precise analysis of curvature and gradients? Recent advances in computer vision and visual psychophysics with natural images cast a doubt on such a naïve hierarchical reconstruction scheme (e.g., Marr, 1982) and propose an alternative approach along the lines of 'task-specific vision' or 'direct perception' (e.g., Gibson, 1979).

Aristotle defined vision as knowing what is where by looking. To achieve this feat, the brain may not be obligated to reconstruct the physical properties (e.g., 3D geometry) of an external object. In principle, the brain could use only the information required to perform a specific task such as recognizing a person and judging her/his attractiveness. Of course, for a biological system, there is a premium on consuming less time and energy as greater efficiency may translate into better odds for natural selection.

One good strategy to achieve task-specific objectives is to use simple image features and their ensemble statistics which are already evident in the retinal image or in early visual cortex. Natural images are far from random—instead, natural images have a particular statistical structure (e.g., 1/f spectrum; e.g., Simoncelli and Olshausen, 2001). Moreover, at the statistical level, natural images often reflect the properties of external objects, scenes, materials, and so on. Thus, if a particular property is reflected in the statistical structure of the image, then the visual system may be capable of estimating that property from it directly. This short-cut strategy needs shallow computation and should therefore be very rapid. One may argue that image-feature representation is too poor for higher-order visual functions. However, recent evidence from psychophysics and computer vision reveals that the human visual system uses image statistics extensively to perform visual judgments on objects, scenes, materials, and their emotional values.

In the field of object recognition, for example, it has been suggested in the late 1990's that the visual system recognizes 3D object based on matching the 2D appearance of an object with canonical viewpoints stored in memory (e.g, Murase and Nayar, 1995). In the early 2000's, several computer vision studies have proposed powerful algorithms for object recognition based on populations of edge features (e.g., SIFT, SURF, and HOG; e.g., Lowe, 1999). These models have viewpoint-dependent characteristics consistent with physiological data from object-sensitive neurons in IT/TE (e.g., Poggio and Edelman, 1990; Logothetis et al., 1994).

In the field of scene perception, psychophysical studies have demonstrated that humans can categorize natural scenes with a latency of less than 100 ms (e.g., Thorpe et al., 1996). Ultra-rapid categorization suggests that scene recognition depends at least partially on low-level feature representation in early visual cortex. Subsequent computer vision studies have supported this notion by showing that a simple model can extract the 'gist' of a scene from the distribution of image features (e.g., Oliva and Torralba, 2001).

The perception of surface properties such as glossiness and translucency has rarely been investigated because of its extreme complexity from a reconstruction-scheme standpoint. However, recent psychophysical and computational evidence shows that the human visual system can estimate surface properties based on very simple image statistics such as histogram skewness (e.g., Motoyoshi et al., 2007; Fleming, 2014). Recent studies further demonstrate that humans can judge the comfortableness/unpleasantness of a surface based on image statistics even faster than they recognize its material class (Motoyoshi and Mori, 2016).

All the aforementioned findings suggest a critical role for low-level image features in high-level visual cognition, including the judgment of emotional value. Of course, there is also significant amount of evidence that points to the limitation of image-based perception. For example, simple histogram statistics can explain (only) 75% of glossiness perception in daily objects (Wiebel et al., 2015). This indicates that, if needed, the visual system uses more elaborate information beyond image features. It seems that the visual system relies mostly on simple image features for quick and easy judgments and uses more information for detailed and accurate inspections. These two visual "modes" operate in parallel and are reminiscent of the classical human information-processing distinction of pre-attentive vs. attentive processes, and system 1 vs system 2 (e.g., Wolfe, 1998; Kahneman, 2003). To date, neither the neural representation nor the inherent computations involved in the latter process are understood. Future investigations may unveil them.

In addition to theoretical achievements, the series of studies cited here send us an important message with respect to psychophysical experiments: Experiments with natural or naturalistic images may lead us to the essence of the problem, whereas experiments using overly simplified stimuli—which are often designed based on a naïve understanding of the visual hierarchy—may mislead us. It should be always kept in mind that biological visual systems have evolved to interact efficiently with the natural world with limited resource and time.

Acknowledgements

This research was partially supported by JSPS KAKENHI JP15H05916 and JP15H03461.

References

Fleming, R. W. (2014). Visual perception of materials and their properties. *Vision Research,* 94, 62–75.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception.* Boston, MA: Houghton Mifflin.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review,* 1449–1475.

Logothetis, N., Pauls, J., Bülthoff, H., & Poggio, T. (1994). Shape representation in the inferior temporal cortex of monkeys. *Current Biology,* 4, 401–414.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. *The Proceedings of the Seventh IEEE International Conference on Computer Vision,* 1150–1157.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* San Francisco, CA: Freeman.

Motoyoshi, I., Nishida, S., Sharan, L., & Adelson, E. H. (2007). Image statistics and the perception of surface qualities. *Nature,* 447, 206–209.

Motoyoshi, I. & Mori, S. (2016). Image statistics and the affective responses to visual surfaces. *Journal of Vision,* 16(12):645.

Murase, H., & Nayar, S. K. (1995). Visual learning and recognition of 3-D objects from appearance. *International Journal of Computer Vision,* 14, 5–24.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision,* 42, 145–175.

Poggio, T., & Edelman, S. A. (1990). Network that learns to recognize 3D objects. *Nature,* 343, 263–266.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience,* 24, 1193–1216.

Thorpe, S., Fize, D., & Marlot., C. (1996). Speed of processing in the human visual system. *Nature,* 381, 520–522.

Wiebel, C. B., Toscani, M., & Gegenfurtner, K. R. (2015). Statistical correlates of perceived gloss in natural images. *Vision Research,* 115, 175–187.

Wolfe, J. M. (1998). Visual Search. In H. E. Pashler (Ed.), *Attention.* East Sussex, UK: Psychology Press.

Part II
Symposium 1—Time in Perception and the Brain

An Introduction to Symposium 1: Time in Perception and the Brain

Symposium Organizer: Makoto Ichikawa

Time is one of fundamental subjects in psychophysics, because the mental time is not the same as the physical time and neural time, and because it is related to anomalous behaviors of the human perception and action and thus helps us to understand the basis for both perception and action. Various approaches are possible to the researches on time in perception. Among them we will concentrate on psychophysical approaches and brain imaging approaches, where the former places emphasis on experimental observation whereas the latter places emphasis on neural basis of the time perception. In this symposium, we invite three researchers from psychophysical research field and one researcher from the brain imaging research field. Makoto Ichikawa will talk how cognitive and emotional factors affect the perception of duration, as well as the temporal precision in visual perception. Vincent Laflamme will show how the perception of the duration of emotional stimulus is affected by sex differences, social context, and attention. Masahiko Terao will show how the visual system integrates retinal inputs over space and time in peripheral vision for perception and saccadic eye movement. Masamichi Hayashi will show how the population coding are involved in the basis of encoding time intervals across parietal and medial premotor cortices by the use of functional fMRI.

DOES THE MISSED FRAME IN VIEWING RSVP DISPLAY AFFECT THE PERCEIVED DURATION?

Makoto Ichikawa
Department of Psychology, Faculty of Humanities Institutet and Department of Psychology, Chiba University, 263-8522 Chiba, Japan
`<michikawa@chiba-u.jp>`

Abstract

It has been known that the perceived duration reduces with the decrease of perceived events during the period. We examined whether failure to detect targets in RSVP (Rapid Serial Visual Presentation) display, which indicates the reduction of perceived frames, causes the reduction of perceived duration by the use of attentional blink paradigm. In each trial, two series of RSVP display were presented; in the first display, two, one, or no numerals were presented as targets within a series of alphabets while, in the second display, only alphabets were presented. We found that failure in target detection caused no reduction in perceived duration. However, the perceived duration for the two-targets-condition with no detection failure was shorter than those for the no-target-condition. These results suggest that the perceived duration in viewing RSVP display reduces not with the decrease of perceived event, but with the increase of cognitive load in target detection.

Perceived duration for a given period decreased with the decrease of perceived event frequency during the period (e.g., Brown, 1995; Fraisse, 1963; Poynter, 1989). In viewing RSVP (Rapid Serial Visual Presentation) display, which includes two targets, observers often fail to detect the second target if the lag between the first and second targets was less than 500 ms. This failure in detecting the second target is known as Attentional Blink (AB). If observers fail to detect either of the targets in viewing RSVP display, they would perceive fewer frames for the RSVP display. In such a case, do they perceive the duration as shorter than the duration of the display for which they detect the targets?

Herbst et al. (2012) reported that the number of subjectively perceived target stimuli determines subjective duration in viewing RSVP display. In this previous study, target was always the same character (X) in RSVP display, which presented only alphabetic characters. In their experiment, the cognitive load could be lite because the target was fixed. In the present study, we examined how failure in target detection task for RSVP display with more cognitive load affects the perceived duration for the display by the use of different alphabetic characters as targets in a series of numerals. In order to vary cognitive load in target detection task, we prepared the No-target condition, in which RSVP display include no target for the detection task, as well as the One-target condition and Two-targets condition, in which RSVP display included one or two targets, respectively.

Experiment

In each trial, two series of RSVP display were presented one after another. Eight participants conducted target detection task for the first RSVP display, and then they conducted relative length judgment task between the first and second RSVP displays.

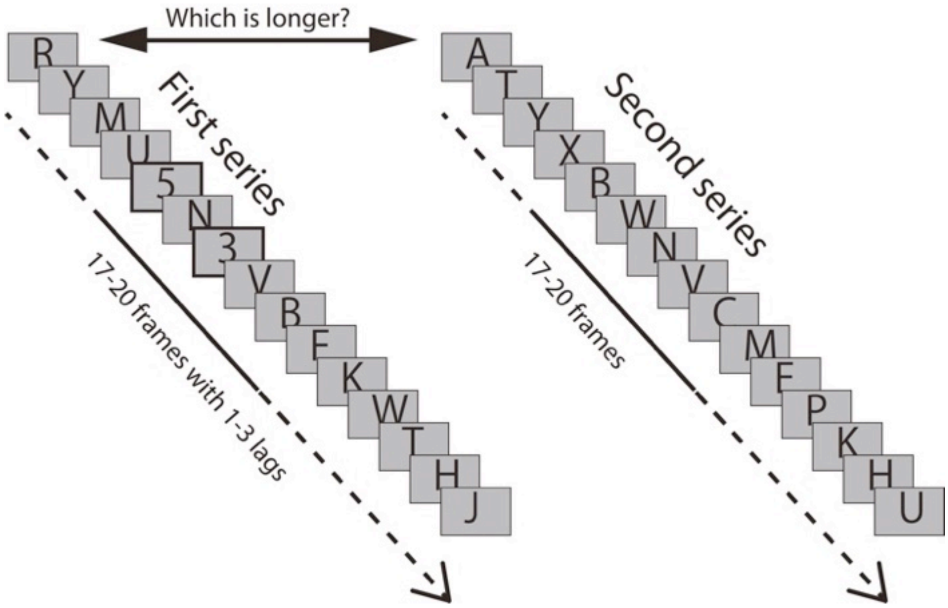


Fig. 1. Diagram of sequence in each trial of the Two targets condition.

Methods

Stimulus. In each trial, two RSVP series with black uppercase alphabets (1.47 × 1.47 deg, 0.03 cd/m^2) were presented on gray background (1.0 cd/m^2) (Fig. 1). That is, the first RSVP series might include black numerals (2, 3, 4, 6, 7, 8, or 9) as targets while the second series presented only alphabets. Each frame was presented for 70 ms with 23 ms of inter stimuli interval. The length of each RSVP series ranged from 17 to 20 frames. Viewing distance was 57 cm. In each trial, the frame difference between the first and second series was either of -1, 0 or +1 (-1 indicates that the first series was shorter than the second one by one frame).

In the Two-targets condition and One-target condition, at least 4 alphabets were presented before the first target. In the Two-targets condition, lag between the first target (T1) and second target (T2) was 1, 2, or 3 frames.

There were three conditions for the first RSVP series. a) Two-targets condition: two numerals were presented as targets within a series of uppercase alphabets [3 (frame difference) × 3 (lag) × 16 (repeat) = 144 trials]. b) One-target condition: one numeral was presented as a target within series of uppercase alphabets [3 (frame difference) × 16 (repeat) = 48 trials]. The serial position of numeral target corresponded to the T1 or T2 position in the Two-targets condition. c) No-target condition: only uppercase alphabets were presented [3 (frame difference) × 16 (repeat) = 48 trials]. Therefore, each observer had 240 trials.

Procedures. Each condition was presented in random order. In each trial, observers pressed the space key to start the first and second RSVP series. After the presentation of the second series, observers reported two target numerals by pressing number keys. They were allowed to press the "miss" key when they did not see target(s). Then, they reported which of the first and second series looked longer.

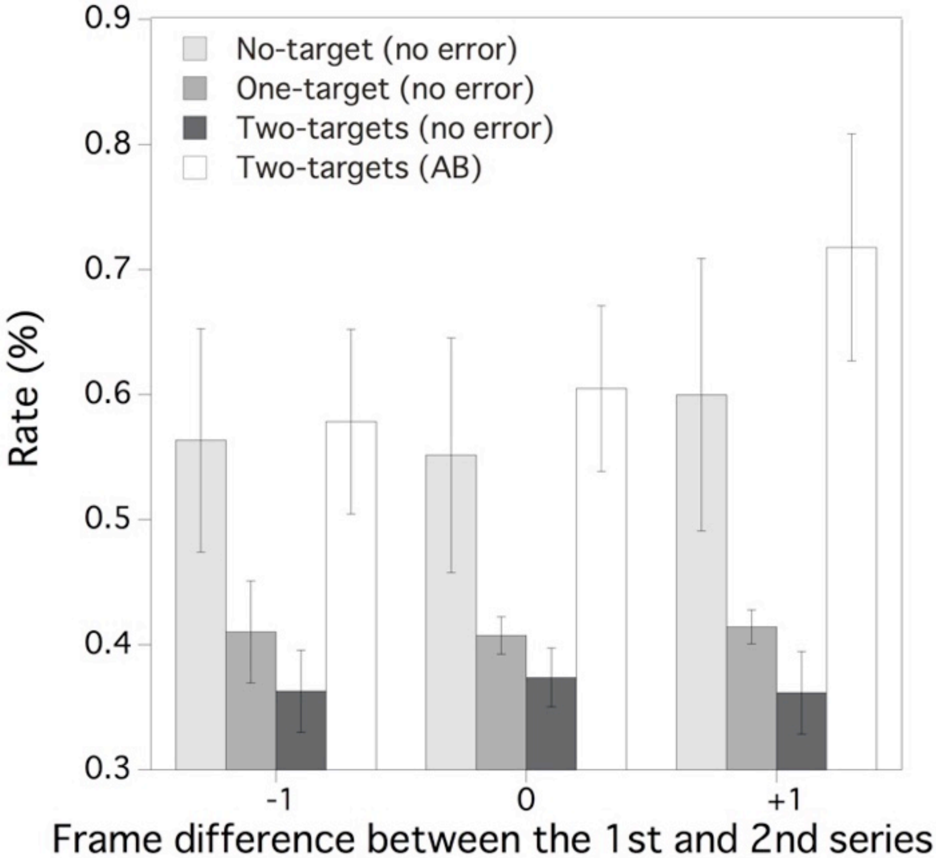


Fig. 2. Rate of duration judgment. Each bar shows the frequency rate of the trials in which the first series was perceived as longer than the second series. Error bars show SEM.

Results and Discussion

For the Two-target condition and One-target condition, if observers reported wrong numerals as target, and if they missed the target, we counted their report as error. Also, if observers reported any numerals as target for the No-target condition, we counted their report as error. Error rates in the One-target and No-target conditions were respectively 9.1% and 5.2% while it was 39.1% for the Two-target condition. For the Two-target condition, frequency of the attentional blink was 24.5%. Frequency of T1 detection error was 0.7%.

We calculated the rate of the trials in which observers reported the first series as longer one for the error trials and no error trials in each condition. By the use of these rates, we conducted a two-way repeated measures analysis of variance (ANOVA) with the frame difference between the series (-1, 0, and +1) and trials (no error trials for the No-target, One-target, and Two-targets conditions, and trials with the attentional blink for the Two-target condition) as factors (Data of one observer who obtained no attentional blink in one of the frame difference condition was eliminated from this analysis). Fig. 2 shows the averages from seven observers, which were used in the ANOVA. Main effect of the trial factor was significant $[F(3, 18) = 6.679, p = 0.0032]$ while the main effect of the frame difference $[F(2, 12) = 1.934, p = 0.1871]$ and interaction of the two factors

$[F(6, 36) = 1.524, p = 0.1984]$ were insignificant. Ryan's post hoc test for the main effect of trials found that the rates for the trials with the attention blink and no error trials in the No-target condition were significantly higher than the rates for the other two cases.

Note that the rate in which the first series was reported as longer one varies among the Two-targets, One-target and No-target conditions even if observers were successful in target detection task. That is, even if observers had no error in target detection, the first series, which may include targets, looked longer than the second series in the No-target condition while it looked shorter than the second series in the Two-target and One-target conditions. These results indicate that the successful target detection has effect to shorten the subjective duration in viewing RSVP display.

If observers failed T2 detection in the Two-target conditions (if the attentional blink occurred), they tended to perceive the first series as longer one. This result suggests that the shortening effect on subjective duration, which observed with successful target detection, would be canceled by attentional blink.

The rates in which the first series was reported as longer one was 61.0% if attentional blink was observed for the Two-targets condition. The rates were respectively 25.0% and 52.0% for T1 detection failure in the Two-targets condition and for target detection failure in the One-target condition. These results suggest that the failure in T2 detection (attentional blink) has unique effect on the subjective duration in observing RSVP display.

We found no effect of the real frame difference between two RSVP series on the relative length judgment. In addition, no result shows that successful target detection causes the elongation of perceived duration in viewing RSVP display, compared to failure of target detection. Rather, failure in T2 detection in the Two-targets condition caused the elongation of the subjective duration. These results were very different from those in Herbst et al. (2012). The difference between two studies would be caused by difference in cognitive load required in target detection. The present results suggest that effects of heavy cognitive load in target detection task upon the subjective duration in viewing RSVP display would be stronger than the effects of objective, or subjective number of stimulus.

Acknowledgements

This study is supported by JSPS Grants-in-Aid for scientific research (#25285197, #26285162). The author thanks M. Miyoshi who conducted experiments and date analyses for this research project.

References

Brown, S. W. (1995). Time, change, and motion: The effects of stimulus movement on temporal perception. *Perception & Psychophysics,* 57, 105–116.

Fraisse, P. (1984). Perception and estimation of time. *Annual Review of Psychology,* 35, 1–37.

Herbst, S. K., van der Meer, E., & Busch, N. A. (2012). Attentional selection dilates perceived duration. *Perception,* 41, 883-900.

Poynter, D. (1989). Judging the duration of time intervals: A process of remembering segments of experience. *Advances in Psychology,* 59, 305–331.

THE EFFECT OF NUMERICAL MAGNITUDES ON THE PERCEIVED DURATION OF MULTI-DIGIT OR SINGLE-DIGIT INTERVALS

Vincent Laflamme and Simon Grondin
Université Laval
<vincent.laflamme.1@ulaval.ca>

There is a wealth of evidence supporting the notion that temporal perception is affected by numerical magnitudes. Relatively little is known about this relation in the context of the timing of empty intervals. This is what we sought to address. Participants were asked to compare the duration of two temporal intervals presented successively and marked by a single digit (filled intervals) or two digits (empty intervals), namely 0, 4 or 8. For one of the experimental groups, the first interval was always marked by the digit 0 and for the second by the digit 4 and the third, by the digit 8. Results show a significant interaction effect on perceived time between the nature of the intervals to-be-timed (filled or empty) and numerical values, but no significant effect involving the group factor (See Fig. 1). Empty and filled intervals seem affected differently by numerical magnitudes. We further argue for the special status of the digit "0" for the temporal-numerical association effect.

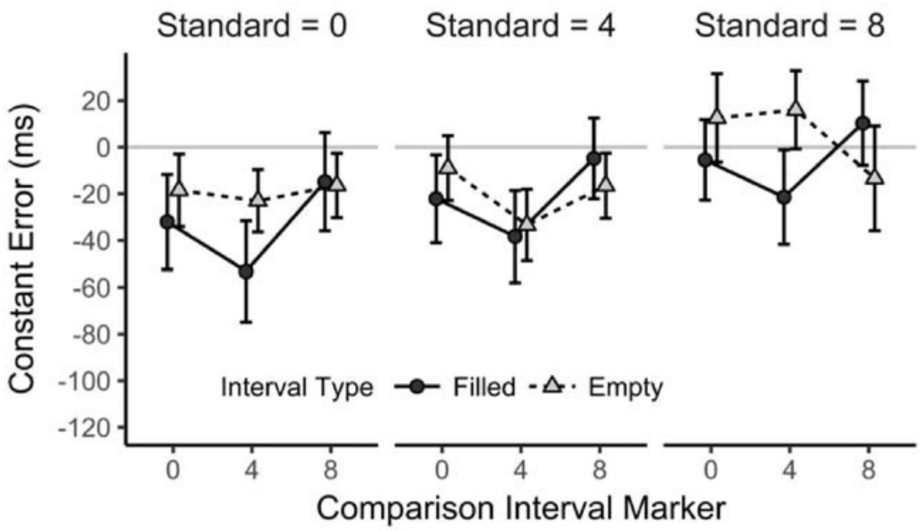


Fig. 1. Mean Constant Error as a function of the Comparison Interval Marker Numerical Value for each Interval Type condition (Filled vs Empty). The panels divide the results according to the different possible values for the marker of the standard interval. Error bars correspond to the standard error of the mean.

VISION AT THE PRESENT MOMENT: VISUAL FEATURE INTEGRATION OVER SPACE AND TIME IN PERIPHERAL VISION

Masahiko Terao
Yamaguchi University, Japan

Psychophysical studies have shown that a retinal input presented later in time backwardly affects percept of another retinal input presented earlier in time. This suggests that visual perception at the present moment is a consequence of pooling visual inputs in the time lag between the timing of a retinal input and the corresponding visual perception. As retinal inputs is highly dynamic and noisy, this pooling of retinal inputs over time may contribute to stable perception. In this talk, I will first introduce a phenomenon in which the appearance of the visual features of an object assimilates to that of the surrounding objects presented later in time in peripheral vision, followed by showing that this backward spatial assimilation rapidly affects saccade control. Finally, I will discuss that this spatio-temporal pooling helps to regularize the appearance among cluttered objects and in turn might overcome the spatio-temporal limitation in peripheral vision.

TIME IN THE BRAIN: NEURONAL CODING OF DURATION

Masamichi J. Hayashi
Department of Psychology, University of California, Berkeley, 94720 Berkeley, USA
Graduate School of Frontier Biosciences, Osaka University, 565-0871 Suita, Japan
<mjhgml@gmail.com>

Time is a fundamental dimension of our perception and action. The current understanding of the neural representations of time is, however, still limited. In this talk, I will show some neurophysiological evidence that duration information is represented by a population code, and this neural representation is associated with our experience of time. First, using neuroimaging techniques we show that repetition of an identical stimulus duration produces a reduction of the blood-oxygenation-level dependent signal in the right inferior parietal lobule (rIPL) (Hayashi et al., 2015). This suggests that the rIPL has a population of neurons tuned for specific durations. A follow-up study further supported this idea by showing that duration information is decodable from the rIPL activity patterns. Finally, we show that the rIPL activity reflects our experience of time rather than simply physical durations of stimuli. Together, our studies demonstrate that subjective experience of time is mediated by duration-tuned neural populations in the rIPL.

References

Hayashi, MJ, Ditye, T, Harada, T, Hashiguchi, M, Sadato, N, Carlson, S, Walsh, V, Kanai, R (2015) Time Adaptation Shows Duration Selectivity in the Human Parietal Cortex. *PLoS Biol,* 13:e1002262.

Part III
Keynote Lecture 2

PROCESSING SPOKEN WORDS IN A SECOND LANGUAGE: COMPETITION, PREDICTION, AND ALIGNMENT

Robert J. Hartsuiker
Department of Experimental Psychology, Ghent University
Henri Dunantlaan 2, B-9000, Ghent, Belgium
<robert.hartsuiker@ugent.be>

Listeners often find it difficult to understand speech in a second language (L2), especially in a noisy environment. Indeed, the unfolding acoustic signal may be compatible with words from multiple languages, complicating lexical access. It is also possible that L2 speech contains phonemes that are absent from one's first language (L1), hindering both perception and production. In my keynote lecture, I will discuss three properties of L2 speech processing that contribute to the difficulty (or ease) of L2 listening: (1) lexical access is language non-selective; (2) processing is predictive; (3) processing is adaptive.

One research line investigates the comprehension of interlingual homophones, words that sound the same but mean something different in the two languages, such as the English word "bay" and the Dutch word "bij" (bee). Auditory lexical decision experiments (both in English and in Dutch) showed that Dutch-English bilinguals respond more slowly to interlingual homophones than to control words. Monolingual English control subjects showed no such effect. The findings suggest that lexical access is language non-selective, so that lexical candidates from both languages compete for selection.

Further studies asked whether listeners can exploit cues to predict upcoming words in L2 (as they do in L1). If L2 listening is more effortful, listeners might find it more difficult to derive such predictions, as predicting might be effortful itself. Eye-tracking studies with English monolinguals demonstrated anticipatory eye-movements towards objects in a visual scene in response to speech. Given "The boy will eat the cake", the number of fixations to the only edible object in the scene (i.e., the cake) will go up right after verb offset (eat) and, importantly, before the object itself is mentioned. We observed that bilingual listeners displayed such prediction effects in both L1 and L2, and to a similar extent. Another lab recently independently reported very similar findings. These studies demonstrate that both L1 and L2 listeners can use lexical-semantic information about verbs to predict upcoming arguments.

A final study investigated whether L2 listeners adapt processing when they listen to a native speaker of that language. In particular, we asked whether exposure to native speech altered their production in L2. As one test case, we considered voiced final consonants in English (as in "pu<u>b</u>") which Dutch-English speakers tend to devoice ("pu<u>p</u>"). There were three experimental phases involving a native English speaker (our confederate) and a native Dutch speaker (the participant). First, the Dutch native speaker read aloud English sentences (baseline). Second, the English native speaker read English sentences (exposure). Third, the two speakers alternated between reading sentences (test). Dutch native speakers shifted towards more voiced productions in the test block vs. the baseline block, suggesting they adapted their speech to that of a native speaker model.

Part IV
Free Talk Session 1

CONVERSATIONAL STYLES OF JAPANESE- AND ENGLISH-SPEAKING CHILDREN AND PARENTS

Yuko Yamashita
Department of English Communication, College of Engineering
Shibaura Institute of Technology, Tokyo, Japan

David Hirsh
School of Education and Social Work, University of Sydney, Sydney, Australia

The present study aimed at exploring the conversational speech style between parents and toddlers from English- and Japanese-speaking families. The study focused on (1) Children's responses toward parents and parents' responses toward children, (2) speech pauses, and (3) lexical use.

In study 1, the responses were classified into three categories: non-lexical backchannels, phrasal backchannels, and repetition. The results showed that the average ratio of overall backchannels and repetitions produced by parents was quite similar in both languages and much greater than that produced by children in the same languages. Among Japanese-speaking parents, non-lexical backchannels and repetitions were preferred to phrasal backchannels, while among English-speaking parents, non-lexical backchannels were most frequently used. With Japanese-speaking parents, almost half of the repetitions were exact in nature. They frequently repeated what a child had said and added the sentence-final particle "ne" or content words.

In study 2, all silent intervals (i.e., pauses) were extracted from each recording. We divided the pause into two types: intra-turn pauses, which occur within parents' or children's utterances, and inter-turn pauses, which occur between the turns of children and parents. Both intra-turn pauses and inter-turn pauses ranged from 0.1 up to 4 s. The results showed that all intra-turn pauses were less than 3 s for Japanese and English parents and 4 s for Japanese and English children. Similarly, all inter-turn pauses at turn transition from children to parents were less than 3 s, and 4 s for all pauses at turn transition from parents to children.

In study 3, all nouns and verbs produced by parents and children were coded and classified into five categories. This study followed the category of nouns and verbs proposed by Choi (1999). Nouns were categorized into object nouns, which referred to animate and inanimate entities, and non-object nouns, which included abstract nouns, locative nouns, and words that described activities or states (Bloom, Tinker & Margulis, 1993; Choi, 1999). Verbs were further categorized into action verbs, stative verbs, and mental verbs. The results showed that nouns and verbs were equally used by Japanese and English parents, while nouns were preferred by Japanese and English children. Additionally, object nouns were more frequently used than non-object nouns, and action verbs were more frequently used than stative and mental verbs in all groups.

These findings are expected to be useful in understanding conversational styles in spoken communication between parents and their children.

AUDITORY OBJECTS REPRESENTED AS TEMPORAL, SPECTRAL, AND SPATIAL MODULATION DENSITY PATTERNS: A THEORETICAL STUDY

Pierre L. Divenyi
Center for Computer Research for Music and Acoustics, Stanford University, USA
<pdivenyi@ccrma.stanford.edu>

Abstract

Experimental studies, e.g., those done at Shamma's and Theunissen's laboratories showing fields of cortical responses to spectro-temporal modulation patterns elicited by meaningful sounds in animals, suggest that similar patterns also underlie auditory object formation by human listeners. The present study expands this representation by combining spectro-temporal modulations with modulations in the spatial dimension, thereby making it possible to completely account for the perception of auditory objects, as defined in the three domains of "what", "when", and "where." Modulation patterns put psychoacoustics on a new footing: Considering objects to be specific modulation patterns in the three domains gives a formal definition to Wundt's concept of aperception; specifying the modulation density along any one dimension reinterprets Fechner's definition of psychophysical discrimination; assessing density changes along all three modulation dimensions combined offers a new way of seeing how and why simultaneous objects perceptually fuse or separate.

A perceptual object was defined by Wilhelm Wundt (Wundt, 1874) as a stimulus that possesses specific values along a set of physical dimensions but one which also represents a recognizable physical object that is meaningful to the person exposed to it. In other words, it should be not only perceived but also *aperceived* by the observer—the object should pass from the sensory-perceptual to the cognitive domain in which it would be learned, recalled, compared to traces of other similar objects, evaluated in terms of its value, meaning, and utility, etc. In particular, auditory objects are important to consider because the acoustic world with its myriads of omnipresent sounds cannot be shut out by humans and by most animals. Simple as they may appear in comparison to the complexity of optical images, sound waves just as these latter ones carry the same basic environmental information regarding the source that emits the stimulus: its identity, its temporal structure, and its location. While these three cardinal properties can be extracted from a waveform, and are decoded by peripheral stages of the binaural auditory system, further processing is necessary to analyze the output of these stages, in order to reconstruct the object emitted by the acoustic source. This more central second stage of analysis can be construed as a process determining the information density along the cardinal dimensions of spectral content, temporal organization, and spatial location. Spectral and temporal information density in the spectral and temporal domains is a result obtained by spectro-temporal modulation transfer functions of peripheral auditory processing in mammals (Kowalski et al., 1996), birds (Singh and Theunissen, 2003), and humans (Elhilali et al., 2003; Elliott and Theunissen, 2009). A recent attempt has been made for extending modulation transfer function analysis to the spatial domain (Divenyi, 2016). The present paper represents an attempt to combine the modulation analysis on all three domains in a theoretical study.

Method

The repertoire of acoustic signals included seven sounds, four of which were synthesized AM signals, one was speech-spectrum noise, and two were sentences taken from publicly available spoken corpora, one by a male talker from the SPIN test list and one by a female talker from the Harvard-IEEE corpus. All signals were 1.7 seconds in length. Each sound was processed to present it at one of 12 azimuthal positions separated by 30 ° over the entire horizontal circle, using the CIPIC head-related transfer function (HRTF) database's KEMAR head recordings with a 2-m radius. From these signals with different azimuthal positions 43 test sounds were generated: eight with single sounds at different azimuths, 12 with two sounds, four with three sounds, and two with four sounds combined. In addition, seven sounds were generated with one or two sounds at two different azimuths and two different AM rates. Peripheral processing of the sounds was emulated (1) by passing them through a bank of ERB-spaced second-order filters followed by an ear-like amplitude compressor, resulting in a cochleagram (Lyon, 1982), and (2) by generating a human-like azimuthal percept based on interaural time difference (ITD) weighted with interaural level difference (ILD), as proposed by Stern and colleagues (Stern et al., 1988), represented both with respect to temporal fluctuations and with the spectral profile of the sound. The modulation processing performed at more central (most likely cortical) sites was implemented by computing two-dimensional Fourier transforms of the three pairings of the three dimensions: time vs. frequency, azimuth vs. time, and azimuth vs. frequency.

Results

The questions that the modulation spectra of the pairs of dimensions allowed to ask were numerous. First, what is the information density along each of the dimensions? To assess the overall magnitude of the density is like evaluating the smoothness of a terrain, for which kurtosis is a reasonable measure. Second, high-intensity areas projected both on the abscissa and on the ordinate of the modulation spectrum reveal at what modulations is the information along the dimension corresponding to the axis is dense. These dense areas may possibly be typical representatives of the auditory object in question. Thirdly, jointly looking at the modulation density of the azimuth and either of the other two dimensions can reveal effects of reverberation (when the azimuthal modulation indicates multiple active angles and when the temporal modulation indicates the presence of a secondary source at a high modulation frequency (i.e., brief delay). Also, analysis of spectro-temporal modulation density can tell if there are two distinct objects or two sounds fusing into a single object—a test of fusion or segregation of two sources, an open question for quite some time (Bregman and Rudnicky, 1975).

Discussion

Clearly, the model as discussed represents the first phase of an in-depth study. Ideally, modulation spectra along the three dimensions would be computed and represented simultaneously in a 3-D drawing. However, one would have to take into account the fact that auditory space is already a three-dimensional concept: azimuth, elevation, and distance of any source viewed from an egocentric point, and such a three-dimensional dimension would make it hard to add it to the dimensions of time and spectrum. Nevertheless, the number of important psychophysical questions that could be treated with such a model may make it worth the effort. Time will tell.

Acknowledgments

The author feels greatly indebted to Drs. Frédéric Theunissen and Richard Stern for discussions about the ideas expressed in this paper.

References

Bregman, A. S., & Rudnicky, A. (1975). Auditory segregation: Stream or streams? *Journal of Experimental Psychology: Human Perception and Performance,* 1, 263–267.

Divenyi, P. L. (2016). Spatial modulation: Hearing the environment. Paper presented at the 22nd International Congress of Acoustics, Buenos Aires.

Elhilali, M., Chi, T., & Shamma, S. (2003). Intelligibility and the spectrotemporal representation of speech in the auditory cortex. *Speech Communication,* 41, 331–348.

Elliott, T. M., & Theunissen, F. E. (2009). The modulation transfer function for speech intelligibility. *PLoS Comput. Biol.,* 5(3), e1000302. 10.1371/journal.pcbi.1000302.

Kowalski, N., Depireux, D. A., & Shamma, S. A. (1996). Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. *J. Neurophysiol.,* 76(5), 3503-3523.

Lyon, R. F. (1982). A computational model of filtering, detection and compression in the cochlea. Paper presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing, Paris.

Singh, N. C., & Theunissen, F. E. (2003). Modulation spectra of natural sounds and ethological theories of auditory processing. *The Journal of the Acoustical Society of America,* 114(6), 3394–3411. `doi:http://dx.doi.org/10.1121/1.1624067`.

Stern, R. M., Zeiberg, A. S., & Trahiotis, C. (1988). Lateralization of complex binaural stimuli: A weighted-image model. *The Journal of the Acoustical Society of America,* 84(1), 156–165.

Wundt, W. M. (1874). *Grundzüge der physiologischen Psychologie* [Fundamentals of physiological psychology]. Leipzig: W. Engelmann.

Part V
Free Talk Session 2

APPARENT TIME COMPRESSION FOR SLOW-MOTION STIMULI

Saya Kashiwakura
Department of Integrated Sciences, The University of Tokyo, Tokyo, Japan
Isamu Motoyoshi
Department of Life Sciences, The University of Tokyo, Tokyo, Japan

A number of psychophysical studies have shown that moving stimuli appear to last longer than static stimuli. Here, we report that this fundamental assumption could be apparently dismissed under specific circumstances. We measured apparent durations of natural movies and artificial drifting gratings presented at various speeds by 2AFC method, using moving stimuli as the comparison. While fast movies were perceived as the longest in accordance with previous studies, slow movies were perceived as shorter compared to static images. However, time compression for slow stimuli disappeared if comparison stimuli were replaced by a white static disk that minimized repetitive exposures to moving stimuli. The results suggest that duration estimation is shaped by the specific distribution of prior visual stimuli to which observers are exposed. A simple model, which includes a rapid recalibration of human time estimation via adaptation to preceding stimuli, succeeds in reproducing our experimental data.

CAUSALITY REPLACES PERCEIVED TIMING OF TWO EVENTS: INVESTIGATION OF THE COMPETITION OF MULTIPLE CAUSAL CONTEXTS IN A SINGLE SEQUENCE

Hiroyuki Umemura
Human Informatics Research Institute, National Institute of Advanced Industrial Science and Technology, 1-1-1 Higashi, Tsukuba, Ibaraki, Japan
<h.umemura@aist.go.jp>

Abstract

A previous study (Umemura, 2017) reported that causal contexts implied in stimuli modulate the perception of the temporal order of two events to accord with the causal contexts. In the stimuli, a ball falling from above and objects moving on the ground were sequentially displayed, and the perceived timings were modulated to accord with the context; a ball hit the ground and then objects moved. The study also reported that the effects of contexts were different among types of motion. It was considered that another context that could imply effects prior to the ball contacting the floor, such as wind pressure caused by the falling ball, competitively affected perceived timings. This suggestion leads to the prediction that this opposite effect would disappear in stimuli that do not comprise a preceding causal context. In the present study, this prediction was tested by using the stimuli that have the same object motions as those of the previous study, but have a different preceding event—the color change of a static object. Even in the display, participants perceived causality between the preceding event and the subsequent motion, e.g., objects were surprised and ran away from the static object that changed its color, while almost all participants did not perceive the preceding causal context. Results showed that the timings of the start of object motion were modulated to accord with the causal context. As expected, shifts of timings were not different among types of motion. This supports the speculation that different effects of causal context on temporal perception among different motions in the previous study would be due to the competitive causal context.

When one sees a ball flying toward a glass and breaking it, one also perceives causality between these two events. In this case, the ball contacting is a 'cause' and the glass breaking is an 'effect'. The perception of causality has attracted many researchers since Michotte showed that simple visual stimuli could give the impression of a causal connection between events (Michotte, 1963). It is relatively recent that studies on the effects of causal perception on other perceptual processes, such as temporal perception, have appeared (Bechlivanidis & Lagnado, 2013, 2016; Buehner, 2012; Eagleman & Holcombe, 2002; Haggard, Clark, & Kalogeras, 2002).

Recently, the author demonstrated that causal contexts implied in stimuli modulate the perception of the temporal order of two events to accord with the causal contexts (Umemura 2017). In the stimuli, a ball falling from above and objects moving on the ground were sequentially displayed. Even though the context was given in subsequent events, the perceived timings were modulated to accord with the context; a ball hit the ground and then objects moved. This effect was strong so that even if the objects on the ground started movement approximately 80 ms before the contact of the ball, participants perceived the movement started after the contact. This study also showed that the effects of causality were different among types of motion. It was speculated that another context, which could have effects on objects before the ball contacted the floor, such as wind pressure caused by a falling ball, competitively affected the modulation of perceived timings. However, there remains a possibility that another cause arising from the difference of motion could affect the modulation of perceived timings.

In the present study, this speculation was tested by using the stimuli that have the same object motions as those of the previous study, but have a different preceding event—the color change of a static object. Although, one might wonder that this display is able to arouse causality perception, observers in the pilot study repeatedly reported the perception of causality. On the other hand, it is difficult to perceive causal effects that are effective before the color change of an object in the present stimuli (the preceding causal context). It is expected that the difference of the shift in temporal order judgement would be reduced with these stimuli.

Method

Apparatus and Stimuli

The experiments were conducted in a dimly lit room. A Windows PC was used to control the stimulus presentation on a CRT monitor placed 40 cm from the observer.

Each stimulus comprised two events. One event was the color change of a cube located at a fixation point ('cause' event). In the other event, ten rings on the floor moved or changed their color ('effect' event). I prepared four types of displays, which were the same as those used in (Umemura 2017). For type A, B, and C stimuli, ten rings moved on the floor in the subsequent event. They are roughly illustrated in Figure 1. For type D stimuli, the rings did not move, but changed their color. The timing of the first event to the start of the second event was chosen from nine inter-stimulus interval (ISI) conditions, -118, -94, -71, -47, -24, 0, 24, 47, and 71 ms, where negative ISIs indicate that a subsequent event started before the change of the cube's color.

Procedure

This experiment included three tasks—temporal order judgement task, causality-rating task, and interview. In the temporal order judgement task, participants judged the temporal order of the change of the cube's color and the start of the rings' movements using a two-alternative forced-choice design. The causality-rating task was conducted after the temporal order judgement tasks were completed. In this task, animations with an ISI of -118, 0, or 71 ms were displayed for participants. Participants chose a higher score (maximum 9) when they strongly perceived causality. All the animations were rated twice. An interview was conducted after the causality-rating task. The same animations were used in the causality-rating task (ISI of -118, 0, or 71 ms). An experimenter showed a participant one of the animations, and asked them to express how the color change of the cube caused the second event if they perceived causality. Participants could freely express the causality they experienced. Eleven observers, oblivious to the experimental purpose, participated in the present study. All the experimental procedures were approved by the Ethics Committee for Human and Animal Research of the National Institute of Advanced Industrial Science and Technology (AIST). All subjects provided written informed consent in accordance with the Declaration of Helsinki before the experiment.

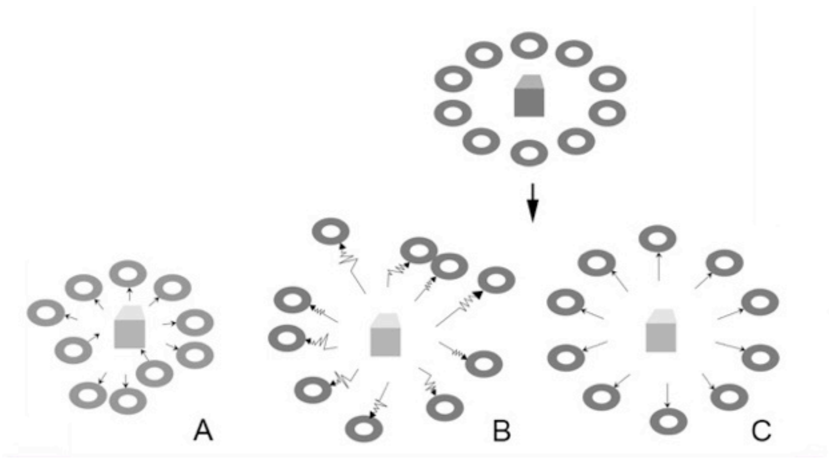


Fig. 1. Illustration of the stimuli. The arrows roughly indicate the directions of movement. D means the color change of the rings. These movements are the same as in (Umemura 2017), and one can watch these movements in the online supplemental materials of the article.

Results and Discussion

I first show the results of causality-rating task and interview here. Table 1 shows the averages of the rating obtained in the causality-rating task. One-way ANOVA for the rating with ISI=0 ms showed a significant effect of the stimulus type ($F(3, 30) = 15.11, p < 0.001$), and a post-hoc comparison with a Bonferroni adjustment revealed significant differences between D and other three displays ($p < 0.001$ in all combinations), but there were no significant difference between any combinations of A, B, and C ($p > 0.5$). Further, the results of ANOVA for the ratings for ISI=-118 ms did not attain the significance level (F(3, 30)=2.33, p=0.09). The ANOVA for ISI=71 ms showed a significant effect of the stimulus type ($F(3, 30) = 5.23, p = 0.005$). A post-hoc comparison showed significant differences between D and A ($p = 0.02$), and D and C ($p = 0.007$), but did not attain a significance level between D and B ($p = 0.06$). No significant effects were revealed between any combinations of A, B, and C ($p > 0.5$).

Table 1. Mean ratings in the causality-rating task and SDs.

Event type	ISI = −181 ms	ISI = 0 ms	ISI = 71 ms
A	2.5 (2.2)	6.5 (2.5)	7.0 (1.6)
B	3.3 (2.6)	7.2 (1.5)	6.8 (1.8)
C	3.4 (2.3)	7.1 (1.5)	7.4 (1.6)
D	1.3 (2.1)	3.2 (2.4)	3.8 (2.5)

The results of the causality-rating indicate that participants perceived causality for this display when ISI=0 or 71 ms. The interpretations for these stimuli acquired through the interview were as follows: "the color of the cube changed and an impulse was generated ($n = 3$)", "the switch was turned on, and then the rings were moved (n=3)," and "the rings were surprised at the change of the cube's color ($n = 2$)." Two participants reported that they did not perceive causality, and one participant felt causality but could

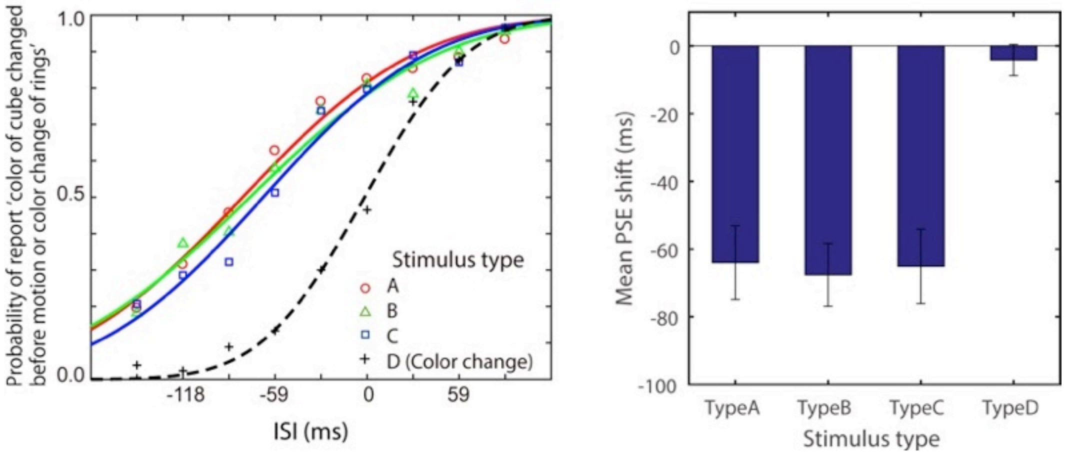


Fig. 2. Results of the temporal order judgment task. (Left) Example of curve fitting. (B) Averaged PSE shift from zero for all participants.

not explain it. Thus, interpretations varied among participants, but did not vary among the types of stimuli. This was different from those of Umemura (2017), in which different interpretations were obtained among different stimuli. More important results for the purpose were that when ISI $= -181$ ms, the ratings were small in three movement types, and almost all participants answered that there were no causal links between the two events.

Figure 2 shows the results of the temporal order judgement task in which participants were required to answer whether the rings started moving before or after the cube changed its color. The proportions of trials in which participants answered that the cube changed its color before the start of the rings' movement (or color change) were plotted and the point of subjective equality (PSE), which indicates that the end of the first event and the start of the second event were simultaneous, was calculated by fitting with a cumulative Gaussian function as an index of the extent to which the timing of the end of the first event and the start of the second event converged (Figure 2, left). The averages of the PSEs are presented on the right of Figure 2. The repeated measures ANOVA revealed a significant main effect of the stimulus type ($F(3, 24) = 31.87, p < 0.001, \eta^2 = 0.629$). A post-hoc comparison with a Bonferroni adjustment revealed a significant difference between D and other three displays ($p < 0.001$ for all pairs), but did not reveal significant differences between any combinations of A, B, and C ($p > 0.5$, for all combinations). This means that the perceived temporal order was transposed to accord with the causal context; the cube changed its color, and at the same time generated impulse. Then the rings started moving, for example. While, there would be no other causal context which differentiate the perceived timing among different motions.

Summary

The results showed that even if the color changing served as a 'cause', participants perceived causality between the event and motions of the rings, and that the timings of the start of the object motion were modulated to accord with the causal context. Almost all participants did not perceive the causal context when the color of the cube changed after the start of the ring's movement. Moreover, as we expected, the shift of timings did not show a significant difference among the three motion conditions. These results support the suggestion that different effects among different motions in the previous study could be due to the competitive causal context. Umemura (2017) suggested that these effects of causal contexts on the temporal order judgement would be explained in the framework of Bayesian estimation, and the difference of effects of the causal context among subsequent object motions would be due to the competition of two hypotheses (priors). Each prior should have a different contribution according to its probability or likelihood. The results of the current study supports this view by showing that equalizing (deleting) the effect of the preceding causal context leads to equalizing the shift of the PSEs.

Lastly, it is interesting that only the changing color of the cube could be a 'cause' for subsequent rings' motion, because I believe we merely encounter such situations. Therefore, we should have more studies on how causal links are acquired and how these links are applied to visual inputs.

References

Bechlivanidis, C., & Lagnado, D. A. (2013). Does the "why" tell us the "when"? *Psychol. Sci.*, 24(8), 1563-1572. doi:10.1177/0956797613476046

Bechlivanidis, C., & Lagnado, D. A. (2016). Time reordered: Causal perception guides the interpretation of temporal order. *Cognition*, 146, 58-66. doi:10.1016/j.cognition.2015.09.001

Buehner, M. J. (2012). Understanding the past, predicting the future: causation, not intentional action, is the root of temporal binding. *Psychol. Sci.*, 23(12), 1490-1497. doi:10.1177/0956797612444612

Eagleman, D. M., & Holcombe, A. O. (2002). Causality and the perception of time. *Trends Cogn. Sci.*, 6(8), 323-325. doi:https://doi.org/10.1016/S1364-6613(02)01945-9

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nat. Neurosci.*, 5(4), 382-385. doi:10.1038/nn827

Michotte, A. E. d. (1963). *The perception of causality* (English translation of Michotte, 1954) (T. R. Miles & E. Miles, Trans.). London: Methuen.

Umemura, H. (2017). Causal Context Presented in Subsequent Event Modifies the Perceived Timing of Cause and Effect, *Front. Psychol.*, 8(314). doi:10.3389/fpsyg.2017.00314

Part VI
Free Talk Session 3

REPRESENTATIONAL MOMENTUM AND ANISOTROPY IN VISUAL SPACE

Timothy L. Hubbard[1] and Susan E. Ruppel[2]
[1]*Arizona State University, Tempe AZ 85281 USA*
[2]*University of South Carolina Upstate, Spartanburg, SC 29303 USA*
<timothyleehubbard@gmail.com>

The possibility of anisotropies in visual space in and near the final location of a moving target was examined. Experiments 1 and 2 presented a moving target, and after the target vanished, participants indicated the final location of the leading or trailing edge of the target. Memory for both edges was displaced forward from the actual final locations, and the magnitude of displacement was smaller for the leading edge. Experiments 3 and 4 also presented stationary objects in front of and behind the final location of the target, and participants indicated the location of the nearest or farthest edge of one of the stationary objects. Memory for the near or far edge of an object in front of the target was displaced backward, and memory for the near or far edge of an object behind the target was displaced forward; the magnitude of displacement was larger for objects in front of the target and when the edge was farther away. The findings (a) suggest representational momentum is associated with an anisotropy of visual space that extends across and outward from the moving target and (b) are consistent with previous findings regarding estimation of time-to-contact, anorthoscopic perception, and memory psychophysics.

EBBINGHAUS ILLUSION IN CONTRAST-DEFINED AND TEXTURE-DEFINED STIMULI

Sofia Lavrenteva and Ikuya Murakami
Department of Psychology, the University of Tokyo, Tokyo, Japan

Shape perception in our vision originates not only from edges defined by the first-order information, i.e., luminance difference, but also from edges defined by the second-order information, such as luminance contrast and texture differences. Here we ask whether shape from these types of information is processed by a common mechanism. To address this issue, we investigated whether a geometrical-optical illusion occurs under conditions of isoluminance when only second-order information is available. We used Ebbinghaus illusion figures that were comprised either of sine-wave gratings or uniform gray disks. The luminance, contrast and grating orientation of the central target were manipulated in order to create stimuli defined by luminance, contrast or texture. The surrounding inducers were always disks filled with uniform gray. We found that the illusion occurred robustly under all of these conditions, which supports the notion of the shared mechanism for shape perception in the first-order and the second-order systems.

INTELLIGIBILITY OF ENGLISH NOISE-VOCODED SPEECH RESYNTHESIZED FROM SPECTRAL-CHANGE FACTORS

Takuya Kishida
Graduate School of Design, Kyushu University, Fukuoka, Japan

Yoshitaka Nakajima, Kazuo Ueda, and Gerard B. Remijn
Department of Human Science/Research Center for Applied Perceptual Science Kyushu University, Fukuoka, Japan

Sophia Arndt and Mark A. Elliott
School of Psychology, National University of Ireland Galway, Galway Republic of Ireland

Previous studies had found that intelligible Japanese speech sounds can be resynthesized from 3 factors extracted with factor analysis of spectral changes of speech sounds [Kishida et al. (2016). Front. Psychol., 7, 517]. These factors consistently appear among different languages [Ueda and Nakajima. (2017). Sci. Rep., 7, 42468]. Here, we investigated whether and how well these spectral-change factors convey linguistic information of British English. We first performed factor analyses of British-English speech. Nine sets of spectral-change factors having different numbers of factors were extracted. Following this, we resynthesized the speech sounds from the extracted factors as noise-vocoded speech. Five men and four women, ranging in age from 20 to 42 years (mean age = 24.4 years, SD = 6.5 years), participated in the intelligibility test of the resynthesized speech. They were all native speakers of Irish English. The participant wrote down what he/she could hear after listening to three repetitions of the speech stimuli. The word identification scores were considerably lower than the mora identification scores in the previous intelligibility test of Japanese noise-vocoded speech. The present performance might have been affected by a higher speech rate. Consistent with the previous test, 3 factors or more were necessary to make the speech stimuli understandable: When the number of factors was 3, the word identification was 23%. The identification performances gradually improved up to 81.6% when the number increased up to 9. The results provide support, although limited, for the importance of 3 or 4 factors in British English. [This study was supported by the JSPS KAKENHI Grant Number 16J05172.]

IRRELEVANT SOUND EFFECTS WITH LOCALLY TIME-REVERSED SPEECH

Wolfgang Ellermeier and Florian Kattner
Technische Universität Darmstadt, Germany

Kazuo Ueda and Yoshitaka Nakajima
Kyushu University, Japan

<[ellermeier, kattner]@psychologie.tu-darmstadt.de
[ueda, nakajima]@design.kyushu-u.ac.jp>

Abstract

Irrelevant sound experiments were performed using three versions of temporally altered speech [Ueda et al. Nature Scientific Reports (2017)]: (1) locally time-reversed speech for which recordings were inverted in segments of 20, 70, or 120 ms duration, (2) reverse playback of thus generated material, and (3) the original recording played backwards. N=38 participants were exposed to these sounds derived from German free-running speech, and another N=43 to the same varieties derived from Japanese—supplemented by control conditions with unaltered speech or noise—while they performed a serial recall task. All were native speakers of German, but did not understand Japanese. The results show significant effects of irrelevant speech on memory which are exacerbated when the utterances become intelligible, namely with the original (German) recording or locally time-reversed speech of short segment duration (20 ms). These effects are attenuated when the distractor is derived from Japanese, unintelligible to our participants. The results are interpreted in terms of phonetic and prosodic processing of the irrelevant sound.

Overhearing irrelevant background speech is known to interfere with the maintenance of information in short-term memory, i.e. the 'irrelevant speech effect' (Banbury, Macken, Tremblay & Jones, 2001; Ellermeier & Zimmer, 2014). To demonstrate that the semantics of the irrelevant stream are not crucial to obtain it, some researchers have used 'reversed speech' (Jones, Miles & Page, 1990; LeCompte, Neely & Wilson, 1997) by playing long utterances from end to beginning, thereby maintaining the overall spectrum and the kinds of spectral changes occurring, but degrading the temporal waveform. Typically, irrelevant speech effects of the same magnitude as with 'forward' speech were obtained; see, however Viswanathan, Dorsi, & George (2014).

Rather than playing an entire speech utterance backwards, the present research focuses on 'locally time-reversed' speech, i.e. temporally inverting short successive segments (20–170 ms long) of an extended speech utterance. Research from our laboratories (Ueda, Nakajima, Ellermeier & Kattner, 2017) assessing the intelligibility of thus locally time-reversed speech had shown it to decrease from perfectly intelligible (at 20 ms segment duration) to practically incomprehensible (> 100 ms), in a very similar fashion for four different languages studied (English, German, Mandarin Chinese, and Japanese).

To investigate, how altering temporal detail in the interfering auditory stream might affect the magnitude of the irrelevant speech effect, we thus used (a) the original recordings, (b) locally time-reversed speech, (c) backward speech, and (d) reversed playback of the entire locally time-reversed sequence—thereby concatenating recovered 'forward' segments in an unnatural order—as distractors in the irrelevant speech paradigm (see Fig. 1 for depictions of these stimulus manipulations). We conjectured that (Hypothesis 1) by degrading 'local phonetics' through time reversal of short segments of the signal, (b) should produce smaller irrelevant speech effects than (d) where the integrity of the speech elements is reinstated (depending, of course, on the duration of the reversed segments). Furthermore, if the large-scale pitch changes in the course of an utterance played a role, we expected that (Hypothesis 2) conditions in which this 'global prosody' is preserved (a and b), would produce greater irrelevant speech effects than conditions in which it is destroyed (c and d). Finally, we wanted to study the effects of our time-reversal manipulations depending on whether the irrelevant stream is derived from the listener's native language, or not.

Method

Participants

Two samples of participants were recruited from the same population: $N = 38$ (10 male, age range 17-43 years, MD = 20) were exposed to German irrelevant speech, $N = 43$ (24 male, age range 18-56 years, MD = 23) to Japanese. The majority consisted of university students participating for course credit; the remainder was paid a honorarium of 8 Euros. All of the participants claimed to have normal hearing.

Apparatus and Stimuli

The speech material (German or Japanese) was extracted from a multilingual database of spoken sentences (NTT-AT, 2002) recorded with a 16-kHz sampling rate and 16-bit quantization. For each trial, some 6-8 sentences were concatenated to produce the desired 14-s irrelevant-speech streams. These were processed as illustrated in Fig. 1: They were either played back as such (original forward speech) or divided up into segments of 20, 70, or 120 ms duration, including 2.5-ms cosine ramps to fade in and out. Subsequently, each segment was reversed in time while maintaining the original order of segments, thus generating 'locally time-reversed' speech streams (depiction (b) in Fig. 1). Additional background speech conditions were produced by 'globally reversing' the thus generated material, i.e. having 'reversed speech' (when the original is played backwards in its entirety), or 'reversed' playback of the previously segmented and locally inverted material ('Rev[ltr]'; see (d) in Fig. 1). For each of the resulting eight 'irrelevant-speech' conditions, 10 different exemplars were generated, thus presenting new acoustical material on each trial.

The stimuli were D/A converted by a high-quality sound card (RME multiface II), subsequently passed through a headphone amplifier (Behringer Pro 8) and diotically delivered to electrodynamic headphones (Beyerdynamic DT 990). Sound levels were adjusted using a 74-dB SPL, 1-kHz calibration tone (supplied with the test sentence database) the level of which was measured at the headphones using an artificial ear (Brüel & Kjær type 4153) fitted with a condenser microphone (Brüel & Kjær type 4192), and connected to a sound level meter (Brüel & Kjær type 2250). The experiment was conducted in a double-walled, sound-attenuated chamber (Industrial Acoustics Company).

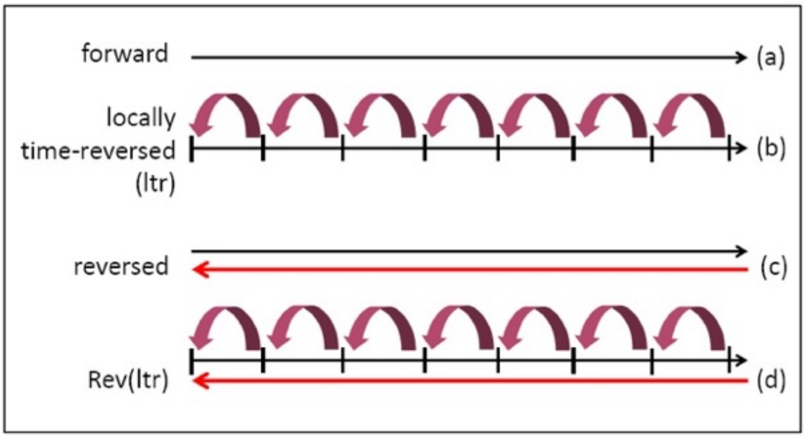


Fig. 1. Schematic representation of the temporal signal processing applied. Top two drawings: Speech utterances were either played forward (as recorded), or locally reversed in time ('ltr'; with segment durations of 20, 70, and 170 ms; these differences not being depicted here). Bottom two drawings: When the entire signal thus obtained was played from end to beginning, reversed speech, or Rev(ltr) conditions resulted.

Procedure

Each trial was initiated by the participant pressing a button. Then, a random permutation of eight digits (drawn from the set of 1 through 9 without repetition) was presented in the center of the screen. Digits were displayed for 1 s each without inter-stimulus intervals. The participants' task was to memorize the order of these digits. After presentation of the digits, a blank screen appeared for 6 s (retention interval) before participants were asked to recall the series of digits. To that effect, a number pad was presented on the screen, and participants had to click on the digits in the respective order. The stream of irrelevant sound (14 s) was played back during the presentation of the digits and during the retention interval. Each background condition (8 derived from speech utterances plus pink noise as a control) was presented 10 times, resulting in a total of 90 trials. The order of trials was randomized for each participant. The experiment proper was preceded by two practice trials, and interrupted by three optional breaks. It took about 60 min to complete.

Results

Performance, i.e. proportion of digits correctly reported at the appropriate position in the sequence, in each of the 9 irrelevant-sound conditions was averaged within and across subjects and is depicted in Fig. 2. The results with German irrelevant speech are depicted in Fig. 2A, the results using Japanese irrelevant speech in Fig. 2B. It is evident that with both background languages, substantial irrelevant speech effects are obtained: Performance in all conditions involving speech (processed or not) is considerably worse than in the pink-noise control conditions (upward-pointing triangles in the two graphs); all $p < .01$ in pairwise Bonferroni-corrected t-tests. Overhearing unprocessed (forward) speech (right-most open circles) appears to constitute the floor for serial-recall performance, and the remaining effects (using temporally processed speech) vary within a range approximately

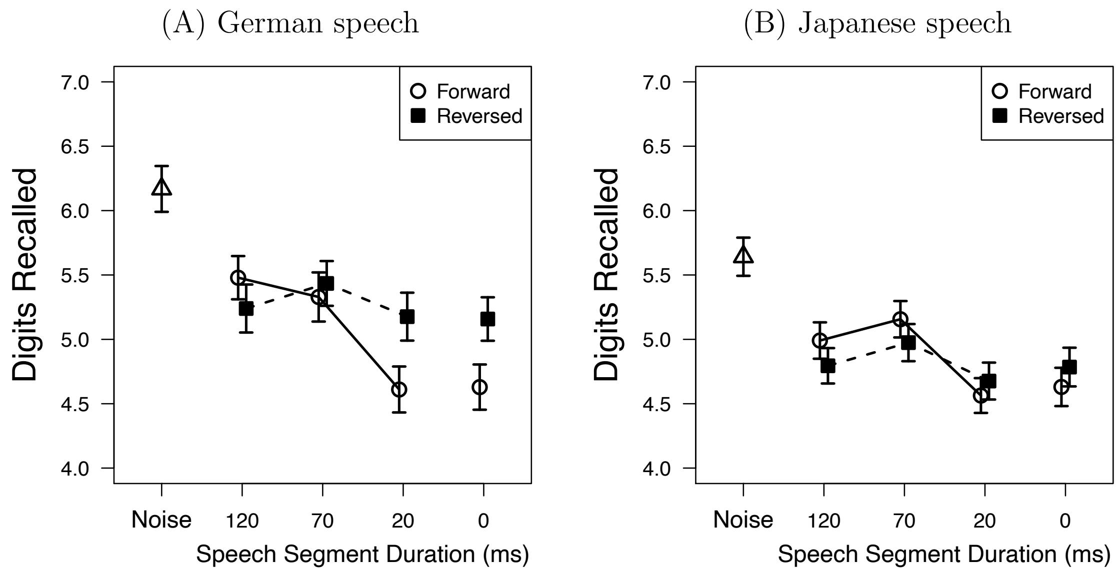


Fig. 2. Performance under irrelevant speech conditions including a 'pink-noise' control condition, 'forward' speech (rightmost open circles in the two graphs), or various conditions derived from locally time-reversed speech (the connected data points), either played as generated (forward) or from end to beginning (reversed). Playing the entire utterance backwards is symbolized by the solitary filled squares at the right of each graph. Means are depicted along with standard errors. The left graph (A) refers to irrelevant sound derived from German speech, the right graph (B) to sound derived from Japanese and presented to $N = 38$ or $N = 43$ German participants, respectively.

covering one-half of the total irrelevant speech effect.

Of primary interest, of course, are the effects of varying segment length and the direction of playback (i.e. ltr and forward speech vs. R(ltr) and globally reversed speech). Therefore, two separate 2 (DIRECTION) × 4 (SEGMENT LENGTH) repeated-measures ANOVAs were performed on the Japanese and German-language data sets, involving all irrelevant speech conditions except for the pink-noise controls. For the German speech material (Fig. 2A), both a significant main effect of SEGMENT LENGTH, $F(3, 111) = 15.11$; $p < .001$, and of reversing DIRECTION emerged, $F(1, 37) = 11.02$; $p < .01$. More importantly, the interaction between the two factors was significant, $F(3, 111) = 8.54$; $p < .001$, suggesting that the effect of segment duration depends on the direction of playback (or that the solid and dashed curves in Fig. 2A are statistically distinct). Post-hoc testing reveals that for the 20-ms segment duration, and for the 'original' (0 ms segment duration), the direction of playback matters ($p = .044$ and $p = .011$, respectively).

The pattern is different when the Japanese-language irrelevant speech effects (observed with German listeners) are analyzed (Fig. 2B): Here, the main effect of SEGMENT LENGTH is statistically significant, $F(3, 126) = 9.88$; $p < .001$, while the main effect of DIRECTION is not $F(1, 42) = 0.21$; $p = .096$; nor is their interaction, $F(3, 126) = 2.16$; $p = .096$, suggesting that the dashed and solid curves in Fig. 2B are statistically indistinguishable. Notably, playing the entire recording forwards or backwards (data points at 0 ms segment duration) makes no difference for the magnitude of the irrelevant speech effect ($p > .99$).

Discussion

The present study shows that sizeable irrelevant speech effects may be obtained with locally time-reversed speech. These effects reach their maximum when the segment duration is short (20 ms), i.e. when the utterances become intelligible (Ueda, Nakajima, Ellermeier & Kattner, 2017). The effect of segment duration, however, is more pronounced in a language the participants understand (German, in the present case). With a foreign language in the irrelevant stream (here: Japanese), the effect of segment duration appears attenuated, as if listeners cannot make sense of the phonetic elements heard in the background. Therefore, successively degrading the local phonetics of the speech stream (by temporally reversing longer and longer signal segments) appears to reduce the irrelevant speech effect to approximately half its original size, and particularly so in a language the listener is accustomed to process. Our initial hypothesis 1, however, that reinstating the integrity of local speech elements, as in 'Rev(ltr)', would produce greater irrelevant speech effects (than 'ltr'), does not appear to be tenable: Rather, at short segment duration, the opposite occurs.

As for hypothesis 2: Reversing the entire sequence (i.e. playing back from end to beginning, see the two bottom depictions in Fig. 1) which was thought to affect the (global) prosody of the utterance, was indeed effective, though only with the familiar language (Fig. 2A). Here, altering global prosody appears to render the percept more 'noise-like', and that is true for the shortest segment duration [Rev(ltr20)] as well as for the entire stream (0 ms segment duration) played backwards.

Clearly, one would like to see further confirmation of these effects by reversing the roles of the two languages studied, i.e. by presenting the same material to Japanese listeners. That is done in a companion paper to appear in the same proceedings.

Acknowledgements

This research constitutes collaborative work jointly performed at TU Darmstadt and Kyushu University. It was partly sponsored by the Japan Society for the Promotion of Science (JSPS). The authors would like to thank Karla Salazar Espino for collecting the German-language data as part of her bachelor thesis, and Maria Hernando for running the Japanese-language conditions as a graduate research assistant.

References

Banbury, S. P., Macken, W. J., Tremblay, S., & Jones, D. M. (2001). Auditory Distraction and Short-Term Memory: Phenomena and Practical Implications. *Human Factors*, 43(1), 12–29. DOI:10.1518/001872001775992462

Ellermeier, W. & Zimmer, K. (2014). The psychoacoustics of the irrelevant sound effect. *Acoustical Science and Technology*, 35(1), 10–16. DOI:10.1250/ast.35.10

Jones, D.M., Miles, C. & Page, J. (1990) Disruption of proofreading by irrelevant speech: Effects of attention, arousal, or memory? *Applied Cognitive Psychology*, 4, 89–108.

LeCompte, D.C., Neely, C.B. & Wilson, J.R. (1997). Irrelevant speech and irrelevant tones: The relative importance of speech to the irrelevant speech effect. *Journal of Experimental Psychology: Learning, Mempory, and Cognition*, 23, 1–12.

NTT-AT (2002). Multilingual speech database 2002 (NTT Advanced Technology Corporation, Tokyo, Japan). URL: http://www.ntt-at.com/product/speech2002/ (date last viewed 6/08/17)

Ueda, K., Nakajima, Y., Ellermeier, W., & Kattner, F. (2017). Intelligibility of locally time-reversed speech: A multilingual comparison. *Scientific Reports*, 7: 1782. doi:10.1038/s41598-017-01831-z

Viswanathan, N., Dorsi, J. & George, S. (2014). The role of speechspecific properties of the background in the irrelevant sound effect. *Quarterly Journal of Experimental Psychology*, 67(3), 581–589.

Part VII

Free Talk Session 4

JOHN SWETS, ROC CURVES, AND FIFTY YEARS OF SIGNAL DETECTION THEORY AND PSYCHOPHYSICS

Roger D. Adams[1], Gordon Waddington[1], Nili Steinberg[2], and Jia Han[3]
[1]*Faculty of Health, University of Canberra, Bruce, ACT 2617, Australia*
[2] *Wingate College of Physical Education and Sport Sciences, Wingate Institute, Netanya, Israel*
[3]*Shanghai University of Sport, Yangpu Qu, Shanghai Shi, China*
<roger.adams@canberra.edu.au; gordon.waddington@canberra.edu.au; knopp@wincol.ac.il; jia.han@canberra.edu.au>

Abstract

One hundred years after Fechner published the Elements of Psychophysics, the Psychometric Society invited involved researchers to contribute commentary papers to a special volume of Psychometrika. In the first article, Edwin Boring argues that Fechner's contribution in founding psychophysics was 'inadvertent'. Two of the papers, by Stanley S Stevens and John A Swets, represent important arms of the psychophysical enterprise, because they define the interests of their authors, in scaling through magnitude estimation and in discrimination through ROC curve analysis, respectively. The death of John Swets in 2016 marks an endpoint that warrants consideration of the half-century of work by the co-author of 'Signal Detection Theory and Psychophysics'. In a current application of ROC analysis in movement psychophysics, the Area Under the Curve is used both as an index of sensory acuity and as an indicator of injury vulnerability, where Youden's Index provides a rule for making decisions about cutoffs.

In 1961, the journal Psychometrika published papers on psychophysics that had been invited presentations for their 1960 Anniversary meeting, held one hundred years after the publication of Fechner's 'Elements of Psychophysics', in 1860. Three of the invitees were Edwin Boring (1886–1968), Stanley S Stevens (1906–1973), and John A Swets (1928–2016). In his article, Boring depicted Fechner as a cavalry commander who attacked the ramparts of materialism, then was inadvertently decorated for measuring sensation. He observed that Stevens' students heard about Fechner, and seldom missed celebrating October 22nd. Stevens questioned how Fechner's concept of error, the jnd, could have become a yardstick for measuring sensation, based as it was on Fechner's erroneous inspiration while lying in his bed, and concluded with his belief in the usefulness of ratio scales of sensation. Swets, however, suggested that Fechner would have welcomed Signal Detection Theory, recognizing in it the ideas about statistical decisions. In his conclusion, Swets (1961) suggested that Signal Detection Theory techniques employed with simple signals, such as operating characteristics, might be applied to more complex areas of research.

Twelve years later, in a paper titled The Relative Operating Characteristic in Psychology, Swets (1973) wrote that the ROC was a technique that effectively isolated the effects of observer response bias in the study of discrimination behavior, and gave a measure of discrimination that was independent of the location of the decision criterion. He noted that although the first appearance of ROC in the literature was in the Tanner and Swets (1954) paper, the Peterson and Birdsall (1953) technical report 'showed us how to plot the data'. In the title of his 1973 paper Swets used the term Relative Operating

Characteristic, though noting that originally, in the detection context, the R stood for Receiver. In his final major paper, Swets, Dawes and Monahan (2000), however, the R in the acronym is again Receiver. In his autobiography, Swets (2010) says ‘the new terminology just didn’t catch on’. Swets (1973) pointed out that if a rating scale is used to obtain better definition of the ROC curve, the cumulative technique employed resulted in a monotonic, increasing curve. Stanislaw and Todorov (1999) state that rating tasks, with r responses, are primarily used to measure sensitivity, and that the area under the ROC curve can be plotted from the r-1 points arising, giving a measure of sensitivity unaffected by response bias. A straightforward interpretation of the ROC area is that it is the proportion of times a subject would identify a signal, if signal and noise were presented simultaneously (Green & Swets, 1966). The area under a ROC curve (AUC) created from confidence ratings can be estimated by application of the trapezoidal rule (Brown, 1974). Because the area of a trapezoid is the multiple of the average length of the two parallel sides by the distance between them, the AUC can be calculated as a sum of a set of trapezoidal areas.

Finally, Swets, Dawes and Monahan (2000) outlined the value of using a non-parametric Signal Detection Theory approach to the problem of obtaining a discriminability measure, and show that not only is the AUC a discrimination measure, but that the ROC curve can be used to determine the best cutoff on a continuous variable to predict a binary state. It is this second application of ROC curves that is the focus of the present study.

When we began to use ROC curves to obtain a discrimination sensitivity measure for the extent of movement at joints (proprioception), we employed five stimuli and five responses, so that the resulting absolute judgement task, by giving the subject a continuous set of numbers to respond with, could be considered to also be a rating scale task. Previous findings have shown the obtained ankle movement discrimination scores, representing proprioceptive sensitivity, to be related to level of athletic achievement (Han et al, 2015), and to be improved by the use of textured insoles (Steinberg et al, 2015) and by training on an unstable balance board (Waddington & Adams, 2004).

The aim of this study was to examine the consequences of a recent ankle injury and to determine the score cutoff for recommending rehabilitation training.

Method

Participants

Forty-two full-time elite classical ballet dancers were in the current study. There were 27 dancers, 13 to 16 years old, and 15 dancers, 16–19 years old. All dancers or their parents provided written informed consent for participation. A survey about any ankle injury in the previous two years was completed prior to participation.

Equipment

Dancers were tested on their non-dominant weight-bearing leg for sensitivity to the extent of active ankle inversion movement on the Active Movement Extent Discrimination Apparatus (AMEDA)

The AMEDA device generates inversion stimuli with small differences in the extent of ankle inversion (Figure 1). With feet shoulder-width apart, the participants stood in a relaxed posture on the AMEDA platform, 50 cm above the floor, with the test foot centred over the axis of the movable base plate. Participants were first given a series of trial movements to perform, to familiarize them with the feel of the five different ankle inversion angles. Each participant performed a total of 15 movements in order to position 1 through 5 in sequence, three times. Participants then undertook 50 non-feedback trials, where they had to respond to the felt ankle inversion position. The trials were in a random sequence of 50, with 10 at each of the five different movement displacements. The stop positions of the stepper motor, operating a moving shaft vertically under the plate, were computer-determined, and ranged between 8 and 12 degrees of inversion, in one degree steps. During the 50 trials, the participants were asked to move down and back to horizontal at a steady pace. After each movement the participant was asked to respond with the number (level) that described the angle at which the device stopped. Software in a laptop computer was used to control a spinning shaft and set the vertical stops for each movement. Thereafter, ROC analysis was used to generate discrimination scores representing each participant's sensitivity to small differences in the extent of ankle inversion.

Data analysis

Data were cast into 5×5 confusion matrices representing the responses made to movements to each of the five stop positions, giving ten trials per row. An example data set is reproduced in the table below.

	Responses				
Stop Position	Response 1	Response 2	Response 3	Response 4	Response 5
1	2	1	4	3	0
2	4	3	3	0	0
3	2	3	2	2	1
4	1	1	2	3	3
5	1	1	1	3	4

The rows were then selected as adjacent pairs, the corresponding ROC curves

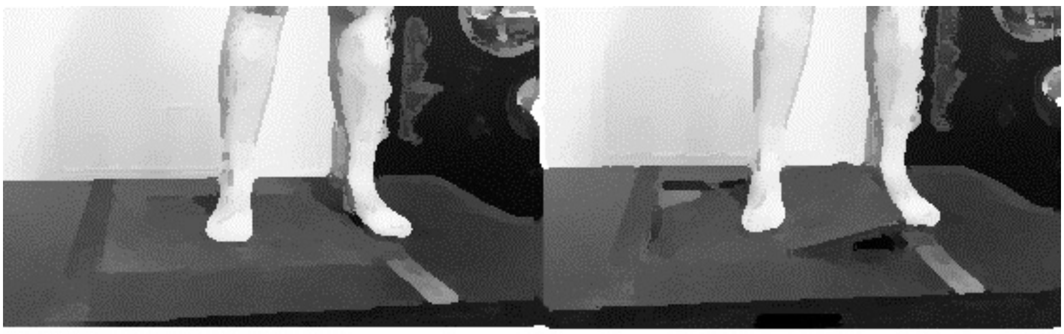


Fig. 1. AMEDA with plate at horizontal, and with ankle inversion to a stop at 10 degrees.

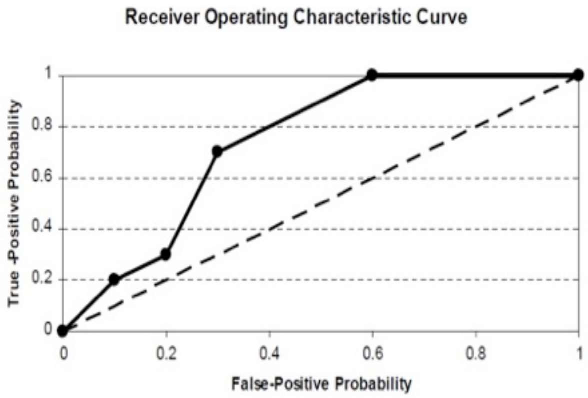


Fig. 2. ROC curve for an adjacent pair of ankle inversion extents. Dropping lines from the points on the curve to the abscissa enables computation of the area in each of the five resulting trapezoids, which when summed and converted to a proportion of the total area give the Area Under the Curve.

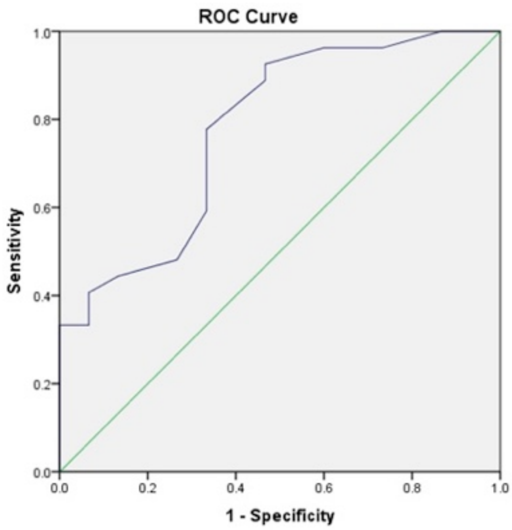


Fig. 3. ROC curve for ankle AMEDA score and Injury/No Injury, with 42 AUC values as the points on the curve, and injury as a 'hit'.

drawn, and the AUC calculated using the trapezoidal rule (Brown, 1974).

Results

Once the four AUC values for each participant were obtained, a mean was calculated to give a single score for ability to discriminate between ankle inversion movements separated by one degree, in the 8 to 12 degrees of ankle inversion range. Of the 42 dancers, 15 reported that they had sprained an ankle in the past two years, and 27 reported no injury. For the group with no ankle injury in the preceding two years, the mean AUC score was M = 0.687 (95% CI 0.668–0.707) whereas for the group with ankle injury the values were M = 0.635 (95% CI 0.608–0.662) and these means were significantly different ($p < 0.01$).

Thereafter, the mean AUC scores as a continuous variable and injury status as a binary variable were entered into ROC analysis. The resulting AUC was 0.778, p = 0.003, 95% CI = 0.629–0.927. Next, Youden's Index was calculated from the Sensitivity and Specificity values, and the AUC value that corresponded to the local maximum identified.

Youden's Index = Sensitivity - (1 - Specificity)

Test Variable: AMEDA_AUC score

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity	Y.I.
.000000	1.000	1.000	0.000
.585000	1.000	.867	0.133
.605000	.963	.733	0.230
.615000	.963	.600	0.363
.625000	.926	.467	0.459
.635000	.889	.467	0.422
.645000	.778	.333	0.445
.655000	.667	.333	0.334
.665000	.593	.333	0.260
.675000	.481	.267	0.214
.685000	.444	.133	0.311
.695000	.407	.067	0.340
.705000	.333	.067	0.266
.715000	.333	.000	0.333
.725000	.259	.000	0.259
.745000	.148	.000	0.148
.765000	.111	.000	0.111
.775000	.074	.000	0.074
.785000	.037	.000	0.037
1.000000	.000	.000	0.000

Discussion

The ROC analysis of the AMEDA ankle inversion discrimination AUC scores of dancers with and without ankle injury in the past two years showed that the AUC score was a significant discriminator between the groups, although at 0.78 it was below the level of 0.8 that Swets gives as the AUC for a test that is a 'good discriminator'. The local maximum in Youden's Index of 0.459, at a mean AUC score of 0.625, provides a cutoff AUC value for differentiating dancers with a previous ankle injury from those with no injury, and by inference a level of ankle inversion discrimination from testing on the AMEDA that warrants ankle rehabilitative work for dancers performing at this level.

Conclusion

The method of ROC curve analysis that John Swets worked to develop provides, in the AUC, a robust measure of discrimination accuracy that has proved to be both sensitive and effective in studies of discrimination of the extent of movements made at different body joints. In particular, at the ankle the measure reflects differences between injured and non-injured ankles, and ankle proprioception differences between athletes with different levels of achievement.

References

Boring, E. G. (1961) Fechner: Inadvertent founder of psychophysics. *Psychometrika,* 26, 3-8.

Brown, J (1974) Recognition assessed by rating and ranking. *British Journal of Psychology,* 65(1), 13-22.

Green, D. M., & Swets, J.A. (1966) *Signal Detection Theory and Psychophysics.* New York: Wiley.

Han, J., Waddington, G., Anson, J., Adams, R. (2015) Level of competitive success achieved by elite athletes and multi-joint proprioceptive ability. *Journal of Science and Medicine in Sport,* 18(1), 77-81.

Peterson, W.W. & Birdsall, T.G. (1953) The Theory of Signal Detectability. *Technical report* No. 13, Electronic Defense Group, University of Michigan, Ann Arbor.

Stanislaw, H. & Todorov, N. (1999) Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, and Computers,* 31(1), 137-149.

Steinberg, N., Waddington, G., Adams, R., Karin, J., Begg, R., Tirosh, O. (2015) Can textured insoles improve ankle proprioception and performance in dancers? *Journal of Sports Sciences,* 34(15), pp. 1430-1437.

Stevens, S. S. (1961) Toward a resolution of the Fechner-Thurstone legacy. *Psychometrika,* 26, 35-47.

Swets, J.A. (1961) Detection theory and psychophysics: A review. *Psychometrika,* 26(1), 49-63.

Swets, J. A. (2010). *Tulips to Thresholds: Counterpoint Careers of the Author and Signal Detection Theory.* Los Altos Hills: Peninsula Publishing.

Swets, J. A. (1973) The Relative Operating Characteristic in Psychology. *Science,* 182, 990-1000.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Better decisions through science. *Scientific American,* October, 82-87.

Tanner W.P. & Swets, J.A. (1954) A decision-making theory of visual detection. *Psychological Review,* 61(6), 401-409.

Waddington, G. & Adams, R. (2004) The effect of a 5-week wobble-board exercise intervention on ability to discriminate different degrees of ankle inversion, barefoot and wearing shoes: A study in healthy elderly. *Journal of the American Geriatrics Society,* 52, 573-576.

GRAY COLOR CATEGORICAL PERCEPTION: ASYMPTOTICS OF PSYCHOMETRIC FUNCTION AND MEAN DECISION TIME

Ren Namae, Marie Watanabe, and Ihor Lubashevsky
University of Aizu
Ikki-machi, Aizu-Wakamatsu, Fukushima 965-8580, Japan
IL: <i-lubash@u-aizu.ac.jp>

Abstract

We summarize the results of our experiments on categorical perception with respect to gray color categorization into two and three classes. Namely, the subjects were instructed to categorize shades of gray (generated in a random sequence) making selection between light-gray and dark-gray (first set of experiments) or between light-gray, gray, and dark-gray (second set of experiments). The collected data are analyzed employing (i) the asymptotics of the constructed psychometric functions and (ii) the mean decision time in categorizing a given gray shade. A plausible macro-level mechanism governing gray color categorization is discussed.

The notion of categorical perception generally describes situations when we perceive our world in terms of discrete categories emerged previously during our communication with the social environment. Our perception is warped such that difference between objects belonging to different categories is accentuated and, in opposite, difference between objects falling into one category is deemphasized (for a review see, e.g., Harnad, 2005; Goldstone and Hendrickson, 2010). As far as color categorization is concerned, in spite of a vast amount of literature about various aspects of color categorization (e.g., Harnad, 2005) the basic mechanisms governing these processes remain up to now a challenging problem.

In the present work we summarize our previous experiments on gray color categorization and shape recognition (Lubashevsky and Watanabe, 2016; Namae et al., 2017) as well as argue for the existence of a certain emergent mechanism of decision making under uncertainty governing the categorical perception at the macro-level rather than the level of particular neurophysiological processes. The pivot point of our experiments is the analysis of (*i*) the *asymptotic behavior* of the corresponding psychometric functions and (*ii*) the mean decision time in classifying a given shade of gray (shade number).

Experimental Setup and Data Processing

On Lenovo LI2221s Monitor (47.7 × 26.8 cm screen) a computer under the operating system Windows 10 visualizes a window of size of 17 × 16 cm with a square $\mathbb{S}$ of size of 11 × 11 cm placed at it center. Color inside this square is changed during experiments; the remaining window part is filled with a neutral gray, namely, RGB(240,240,240). The brightness and contrast of the screen was set equal to 70% and 60%, respectively. To get subject's response to a visualized color we used a standard game joystick. The same computer was used for all the experiments. Integrally eight subjects (five male and three female students) were involved in these experiments.

Each trial of shade categorization is implemented as follows. A random integer $I \in [0, 255]$ is generated uniformly and the area $\mathbb{S}$ is filled with the gray color $G(I) = RGB(I, I, I)$. Then, in one set of experiments, subjects have to classify the visualized gray color $G(I)$ into two categories: *light gray* and *dark gray*, within the other set of experiments into three categories: *light gray*, *gray*, and *dark gray*.

A subject's choice is recorded via pressing the corresponding buttons of a standard joystick. Then a mosaic pattern of various shades of gray is visualized for 0.5 s. This mosaic pattern is used to depress a possible interference between color perception in successive trials that can be caused by human iconic memory. Figure 1 illustrates this. After that a new number I is generated uniformly and the next trial starts.

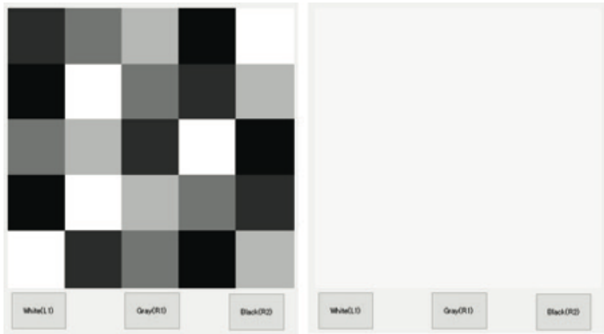


Fig. 1. Color generator

Each data point comprises three quantities (I, J, Δ): the currently generated index I of gray shade, the index J of currently chosen category, and the decision time Δ, i.e., the time interval between visualizing the gray color $G(I)$ and the instant when a subject presses the corresponding joystick button.

Experiments for each subject spanned over 4 days. We have recorded 2000 data points per day and the total amount of data points for one subject individually is $M = 8000$. This amount of data is crucial because, on one hand, only the asymptotic behavior of psychometric functions bears the information about possible mechanisms governing subject's decision-making. On the other hand, the relative volume of collected data related to this asymptotics is rather small.

The collected data have been used for constructing the following two functions. The first one is the psychometric function for a given gray color category J_α, e.g., the "light-gray" category, i.e., the probability $P_w(I)$ of classifying a given shade I as "light-gray"

$$P_\alpha(I) = \left[\sum_{k=1}^{M} \delta(I, I_k)\delta(J_\alpha, J_k)\right] \cdot \left[\sum_{k=1}^{M} \delta(I, I_k)\right]^{-1}, \tag{1}$$

where k is the index of recorded data point and $\delta(i, j)$ is the Kronecker delta. The second function $T(I)$ is the mean decision time of choosing any one of the possible categories for a given shade of gray:

$$T(I) = \left[\sum_{k=1}^{M} \delta(I, I_k)\Delta_k\right] \cdot \left[\sum_{k=1}^{M} \delta(I, I_k)\right]^{-1}. \tag{2}$$

The two functions are used to single out the characteristic properties of plausible mechanisms governing the analyzed categorical perception.

Results and Discussion

Figures 2 and 3 illustrate the obtained results. The logarithmic scale of the P(i.e., y)-axis is used to visualize the asymptotic behavior of the constructed psychometric functions $\{P_\alpha(I)\}$ *vs* the shade number I of gray color. The frames are arranged in such a way that the psychometric function plot and the mean decision time plot combined together for each subject individually be placed one above the other. This arrangement makes it clear that the peaks in the dependence of mean decision time on the shade index correspond to the regions where the uncertainty in gray color categorization is most pronounced.

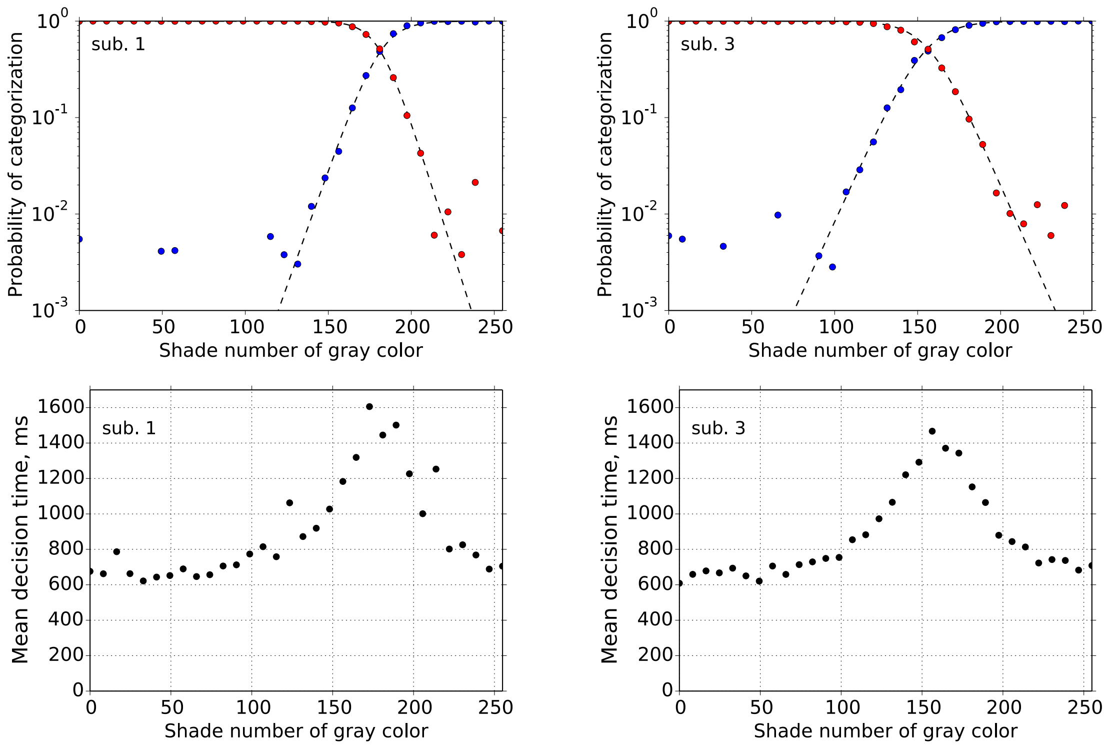


Fig. 2. Psychometric functions $P\alpha(I)$ (or $1 - P\alpha$) and the mean decision time $T(I)$ vs the shade of gray G(I) for two subjects involved in the two-categories-choice experiments.

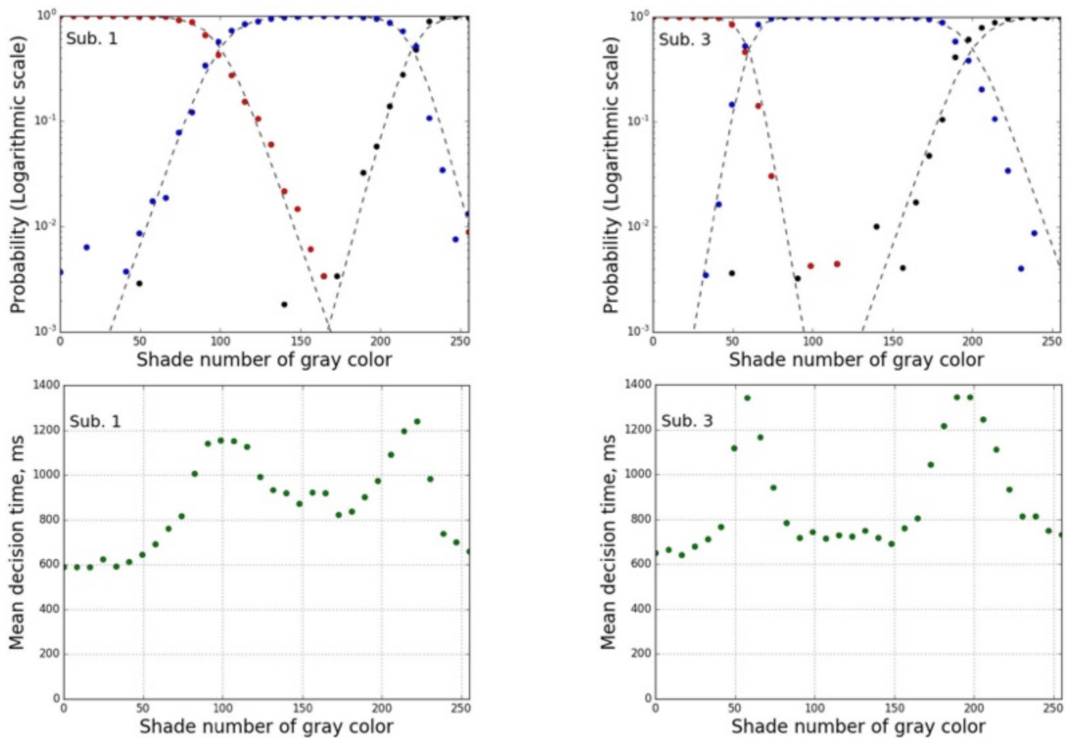


Fig. 3. Psychometric functions $P\alpha(I)$ (or $1 - P\alpha$) and the mean decision time $T(I)$ vs the shade of gray G(I) for two subjects involved in the three-categories-choice experiments.

As seen, there are three common features exhibited by the data collected for all the subjects in the presented experiments.

First, it is the exponential form of their asymptotic behavior. In the shown log-normal plots dashed lines represent the fitting logistic-type functions specified by the fitting function:

$$P_\alpha(I) \approx \frac{1}{2}\left\{1 + \tanh\left[\pm\frac{(I - I_{m:\alpha})}{\delta I_\alpha}\right]\right\}, \tag{3}$$

where $I_{m:\alpha}$ is the center point of the crossover region and the value δI_α characterizes its thickness. The used parameters of this approximation in fitting the experimental data are individual for each subject.

These results allow us to suppose that the mechanism governing the uncertainty in subject's decision-making in categorizing the gray shades admits interpretation as the potential type mechanism. It means that the subject's decision-making is treated as a random process ζ (where ζ is the state variable of some system) in a potential field $U(\zeta, I)$ depending on the shade index I as a parameter. In this case the subject's choice of a category α at trial k for a given shade index I is described by the steady-state probability of finding the system inside the potential well located at the point ζ_α:

$$P_\alpha \approx \exp\{-U_\alpha(I)\}\left[\sum_{\alpha'}\exp\{-U_{\alpha'}\}\right]^{-1} \tag{4}$$

Here $U_\alpha(I) := U(\zeta, I)|_{\zeta=\zeta_\alpha}$ and we have assumed all the potential wells to be rather deep to ignore the contribution of other system states to probability (4). In the general case when for a certain value of the grade index I the contributions of two potential wells become comparable, the other potential wells remain ignorable. So within the linear approximation of the potential dependence on the parameter I

$$U(\zeta, I + \delta I) \approx U(\zeta, I) + \frac{\partial U(\zeta, I)}{\partial I}\cdot \delta I$$

expression (4) leads to formula (3). Other models like detecting noisy signals give another type asymptotics of psychometric functions.

Second, the presented plots show how the mean time required for the subjects to make decision about classifying a current shade of gray changes with the shade number I. For the values of I corresponding to the crossover regions the patterns $T(I)$ contain peaks. In particular, for all the subjects the mean decision time is found to be about 600–700 ms outside these peaks, which can be regarded as the upper boundary T_p of the human response delay time controlled by physiological processes of recognizing threshold events within their unpredictable appearance. At least, it is the upper boundary of visual time intervals presenting timescales relevant to natural behavior, see, e.g., Mayo and Sommer (2013) and references therein.

In the crossover regions the uncertainly in subject's category choice is most essential because for such values of shade number choosing one of two categories becomes equiprobable. Here the peak values T_m of the mean decision time exceed T_p substantially, e.g., for the shown data $T_m \approx 1.4$ s. It should be also noted that a similar dependence of the decision time was found in the speech recognition Bidelman et al. (2013), however, the time delay attained in the peak maximum does not exceed 600 ms. In our data the mean time of decision-making includes also a time interval between making decision and pressing the corresponding button. However, if the time delay is deducted from the measured time data, the found peak will be even more pronounced.

The appearance of these peaks at the $T(I)$-patterns can be explained by turning to the concept of dual system of decision-making. This concept accepts the existence of the automatic and intentional systems contributing simultaneously to human response and being in a continuous interplay with each other. The former—automatic system—is reflexive, fast, affective, associative, and primitive. The latter—intentional system—is deliberative, controlled, slow, cognitive, propositional, and more uniquely human. Besides, there are accounts assuming the dual-processes to arise parallel and compete with each other. However, there are also arguments against the dual system of decision-making; for a review and discussion of the evidence supporting both sides of debate a reader may be referred, e.g., to Rustichini (2008); Evans (2008, 2011). The found peaks of the mean decision time dependence on the shade number argue for that the two cognitive systems do exist and are comparable in their influence on the categorical perception. Indeed, the characteristic time scale of decision-making in categorical perception depends substantially on the uncertainty in the subject's choice. It argues that conscious (mental) processes should be involved in decision-making when the choice of the appropriate color category is not evident.

Third, in categorizing any given shade of gray the subjects made choice only between two classes. Indeed, in the crossover regions the data points correspond to the subject's choice either between "light-gray" and "gray" or between "gray" and "dark-gray." For example, the crossover region of the choice between "light-gray" and "gray" practically does not contain the data points matching the "dark-gray" category.

Conclusion

The reported results argue for the following hypothesis:

- Categorical perception, at least, of shades of gray, is governed by a potential mechanism of decision-making that can be treated as a random process in a potential field whose profile depends on a given shade number as a parameter. Within this approach the classification classes match the wells of the corresponding potential relief.

- The characteristic time scale of decision-making in categorizing the shades of gray, at least in the analyzed case, depends substantially on the uncertainty in classifying a given shade; the higher the uncertainty, the longer the decision time. The obtained data enable us to relate this effect with considerable contribution of mental processes to categorization. We regard this feature as a certain argument for the existence of the dual system of cognitive processes.

- When it is physically possible and different categories are independent, i.e., they are not mixed in the mind—which is the case in the gray color categorization—humans prefer to make choice between only two categories of classification for a given object or stimulus, whereas other categories are not taken into account. We call it the principle of pair-wise categorization.

References

Bidelman, G. M., Moreno, S. and Alain, C.: 2013, Tracing the emergence of categorical speech perception in the human auditory system, *NeuroImage* **79**, 201–212.

Evans, J.: 2008, Dual-processing accounts of reasoning, judgment, and social cognition, *Annual Review of Psychology* **59**, 255–278.

Evans, J. S.: 2011, Dual-process theories of reasoning: Contemporary issues and developmental applications, *Developmental Review* **31**(2–3), 86–102. Special Issue: Dual-Process Theories of Cognitive Development.

Goldstone, R. L. and Hendrickson, A. T.: 2010, Categorical perception, *Wiley Interdisciplinary Reviews: Cognitive Science* **1**(1), 69–78.

Harnad, S.: 2005, To Cognize is to Categorize: Cognition is Categorization, *in* H. Cohen and C. Lefebvre (eds), *Handbook of Categorization in Cognitive Science*, Elsevier, Amsterdam, pp. 20–43.

Lubashevsky, I. and Watanabe, M.: 2016, Statistical Properties of Gray Color Categorization: Asymptotics of Psychometric Function, *Proceedings of the 47th ISCIE International Symposium on Stochastic Systems Theory and Its Applications Honolulu, Dec. 5-8, 2015*, Institute of Systems, Control and Information Engineers (ISCIE), Kyoto, pp. 41–49.

Mayo, J. P. and Sommer, M. A.: 2013, Neuronal correlates of visual time perception at brief timescales, *Proceedings of the National Academy of Sciences* **110**(4), 1506–1511.

Namae, R., Watanabe, M. and Lubashevsky, I.: 2017, Gray Color Multi-Categorical Perception: Asymptotics of Psychometric Function, *Proceedings of the 48th ISCIE International Symposium on Stochastic Systems Theory and Its Applications Fukuoka, Nov. 4-5, 2016*, Institute of Systems, Control and Information Engineers (ISCIE), Kyoto, pp. 76–80.

Rustichini, A.: 2008, Dual or unitary system? Two alternative models of decision making, *Cognitive, Affective, & Behavioral Neuroscience* **8**(4), 355–362.

SUBJECTIVE CONFIDENCE IN PERCEPTUAL CATEGORIZATION: ALTERING PERFORMANCE ASYMPTOTES INCREASES OVERCONFIDENCE

Jordan Richard Schoenherr and Guy Lacroix
Department of Psychology, Carleton University
1125 Colonel By Drive, Ottawa, ON K1S5B6 Canada
<Jordan.Schoenherr@Carleton.ca, Guy.Lacroix@Carleton.ca>

Abstract

Categorization represents the coordination of perceptual discrimination and memory processes. Contemporary models have acknowledged the conjoint contributions of implicit and explicit processes, notably in COVIS (competition between verbal and implicit systems). COVIS assumes that an explicit system engages in hypothesis-testing, dominating the early stages of learning. Over time an implicit system that uses feedback to engage in procedural learning, dominating later stages of learning. Studies supporting this theory have focused on categorization responses and concurrent tasks. Using the randomization technique, we presented participants with Gabor patches from two contrasting categories. We varied the extent to which exemplars from the categories overlapped, creating a performance asymptote (65%) allowing us to examine overconfidence. Participants required little training to reach this performance asymptote Importantly, in contrast to conditions with higher performance asymptotes (85%), we observed much greater overconfidence. This suggests that confidence reports are not solely or primarily determined by accumulated evidence.

The acquisition, development, and use of categories have been an enduring topic of study across psychology. These studies have yielded the progressively clearer understanding that category learning relies on two interactive, yet dissociable information processing systems. The first is a fast-learning, resource-limited explicit hypothesis-testing system that can learn in the absence of feedback. The second system is a slow-learning, high-capacity implicit procedural-learning system that is feedback-dependent (e.g., Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Erickson & Krushke, 1998; Nosofsky, Palmeri & McKinley, 1994). During learning, a representational shift occurs from early to late stages of learning: early stages of categorization are dominated by an explicit representational system whereas the later stages of learning are dominated by an implicit representational system (e.g., Ashby et al., 1998).

A straightforward implication of dual-process models of categorization such as COVIS (COmpetition between Verbal & Implicit Systems; Ashby et al., 1998) is that participants should have more explicit awareness of rule-based category structures relative to information-integration category structures. However, few studies have examined the role of subjective awareness (Paul et al., 2011) or have attempted to dissociate these learning systems with subjective measures in categorization (Schoenherr & Lacroix, 2012). The present study examines whether differences in subjective awareness are evidenced over the course of learning and whether different category structures are associated with greater reported levels of subjective awareness. Specifically, we employed post-decisional confidence reports to assess subjective confidence and compared these results with participants' accuracy (i.e., confidence calibration). By examining systematic variations in overconfidence, we provide evidence for changes in subjective awareness.

Competition between Verbal and Implicit Systems

COVIS assumes that participants make categorization decisions using a criterion or category boundary (e.g., Ashby & Gott, 1988). When participants are presented with a stimulus, they categorize it relative to the category boundary. With feedback, participants will adjust the location of the category boundary to maximize the separation between the categories. In the hypothesis-testing system, simple low-dimensional rules (e.g., stimulus size or frequency) are rapidly generated and tested. In the procedural-learning system, higher-dimensional rules (e.g., stimulus size and frequency) are slowly created and altered with feedback. Ashby et al. (1998) additionally assumed that the hypothesis-testing system and the procedural-learning system remain co-activated under most conditions. In this way, they compete for response selection with the hypothesis-testing system dominating early stages of training and the procedural-learning systems dominating later stages of training. While evidence supporting these results comes from an examination of response accuracy, we sought out a subjective measure of awareness that could be used throughout learning.

Confidence Reports and Confidence Processing

Subjective measures of performance such as confidence reports have a long history in psychophysical research (for a review, see Baranski & Petrusic, 1998). In most studies, participants provide a post-decisional confidence reports by assigning a subjective probability on a rating scale immediately after they have made a decision. For instance, in a two-alternative forced-choice (2AFC) task, participants might indicate that they were guessing (50%), that they had a reasonable level of confidence (e.g., 70%), or that they were certain in their response (100%). 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., 0.7) and mean confidence are equivalent (e.g., 70%) whereas miscalibration such as overconfidence 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.

COVIS contains specific assumptions concerning confidence. It assumes that confidence is obtained by a direct-scaling of this evidence obtained during the primary decision (Ashby et al., 1998). This conforms to SDT-based models of confidence process that claim that information obtained from the primary decision is the sole determinant of confidence report (e.g., Ferrel & McGooey, 1980). However, a number of studies have suggested that confidence reports are affected by sources of information other than that provided by the target stimulus (Busey, Tunnicliff, Loftus, & Loftus, 2000; Schoenherr, Leth-Steensen, & Petrusic, 2010) and require additional operations associated with increased decision response time (DRT) when they are reported (e.g., Baranski & Petrusic, 1998; Schoenherr, 2009). Direct-scaling models have difficulty accounting for such findings (cf. Pleskac & Busemeyer, 2010). Subjective awareness of the properties different category structures should influence participants confidence, with greater awareness of a categorical representation associated with greater confidence. On this account, overconfidence could suggest that while participants can access a representation of the category structure, this representation is not what is being used to categorize stimuli. We consider the basis for these predictions next.

Subjective Awareness in Category Learning

Schoenherr and Lacroix (2012) have explored the possibility of using confidence reports to dissociate category learning systems. They assumed 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 an asymptote prior to accuracy given the flexibility of the hypothesis-testing system and should exhibit a more rapid learning rate. Second, they assumed 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.

Schoenherr and Lacroix (2012) obtained evidence across three experiments that supported dissociable category learning systems. Using an 85% performance asymptote (Experiments 1a and 1b) while also manipulating feedback in order to affect learning in the procedural-learning system but not in the hypothesis-testing system (Experiment 2), they observed increases in DRT when post-decisional confidence reports were required in comparison to a no confidence condition (Experiment 1a). They additionally observed increased overconfidence in intermediate phases of training for those participants learning a rule-based category structure relative to those who learned the information-integration category structure. The pattern of miscalibration suggested that the representation used to report subjective confidence were influenced by different sources of information. They interpreted the greater level of overconfidence as evidence for greater subjective awareness of rule-based representations that failed to contain information concerning the stimulus variability (i.e., they ignored exception exemplars).

In the present study, we sought to extend our previous findings. All manipulations were identical to those of Schoenherr and Lacroix (2012) except that a category overlap of 35% was used (See Ell & Ashby, 2006). The resulting performance asymptote of 65% was imposed to increase the proportion of negative feedback that participants received relative to Schoenherr and Lacroix (2012). We thus predicted that participants would exhibit greater overconfidence relative to our previous experiments if participants failed to account for the increase in negative feedback or exception exemplars.

Method

The 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 65% 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.

Results and Discussion

We first examined whether any participants performed at or below chance using a one-sample t-test. It revealed that all participants' classification accuracy were above chance ($M = .660, SD = .071$) suggesting that participants had learned the category structures, $t(102) = 22.86, p < .001$. Thus, no participants were excluded in either the rule-based or information-integration conditions.

Proportion Correct

A 2 (Categorization Rule: rule-based vs. information-integration) × 2 (Confidence Condition: block-level confidence vs. trial-and-block confidence) × 12 (Experimental Blocks: 1-12) mixed-design ANOVA was performed on proportion correct, collapsing across response key. Replicating previous studies using similar category structures (Ell & Ashby, 2006), no interaction was observed between Categorization Rule and Experimental Block, $F(11, 924) = .92, p = .504$. Replicating the results of our previous experiments (Schoenherr & Lacroix, 2012), we observed a significant effect of Experimental Block, $F(11, 924) = 10.21, MSE = .01, p < .001, {\eta^2}_p = .11$, as well as the main effect of Categorization Rule, $F(1, 84) = 4.15, MSE = .06, p = .045, {\eta^2}_p = .05$. Neither the main effect of confidence condition, $F(1, 84) = .099, p = .754$, nor its interactions with Experimental Block or Categorization Rule were significant (all $Fs < .91$, all $ps > .50$).

Table 1. Averaged dependent variables for rule-based (1D) and information-integration (2D) category structures. Standard errors are presented in parentheses.

	$p(correct)$	M_{Conf}	Calibration	O/U
Rule-Based	.64 (.01)	81.96 (2.36)	.07 (.01)	.16 (.02)
Information-Integration	.67 (.01)	78.27 (2.17)	.05 (.01)	.09 (.02)

As in Schoenherr and Lacroix (2012), participants' accuracy increased over blocks of learning trials. In the present study, we observed that learning was slight better in the information-integration condition relative to the rule-based condition (see Table 1). However, the mean performance (M = .66) across both conditions was nearly identical to the desired performance asymptote (i.e., a learning criterion of .65).

Trial-Level Subjective Confidence Calibration

A 2 (Categorization Rule: rule-based vs. information-integration) × 4 (Experimental Phase: 1-4) repeated-measures ANOVA was conducted. An analysis of subjective calibration did not reveal any difference between Experimental Phases, $F(3, 132) = 2.088, p = .12$, and obtained only a marginal effect of Categorization Rule (see Table 1), $F(1, 44) = 2.83, MSE = .00, p = .10$. The interaction of Experimental Phase and Categorization Rule was also not found to be significant, $F(3, 132) = .428, p = .696$. Such

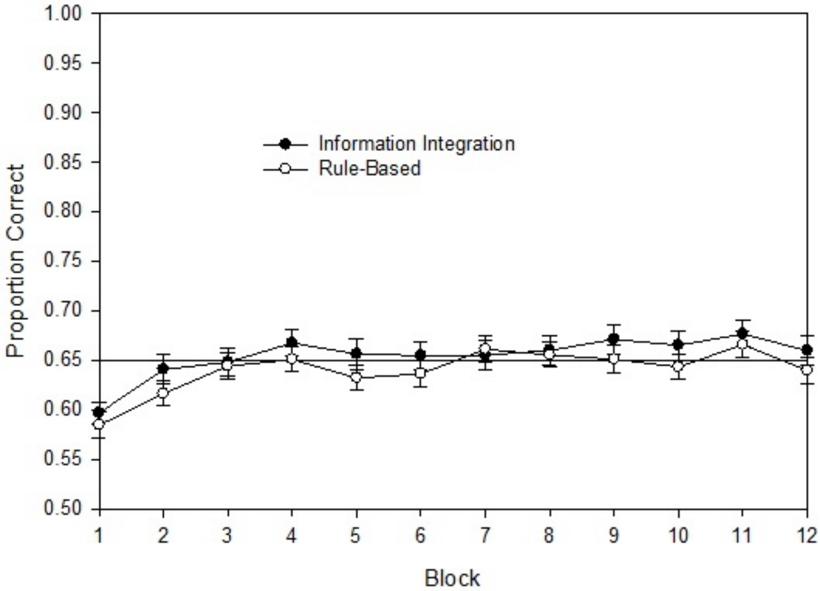


Fig. 1. Response accuracy with 65% performance asymptote. The error bars represent the standard error of the mean.

a finding is expected given that both accuracy and mean confidence reports reached an asymptote early in training. It suggests that, on a trial-to-trial basis, participants had a similar level of awareness of their performance although there is a trend toward improved calibration from the early phases of training to later experimental phases.

Trial-Level Overconfidence Bias

Unlike the calibration analysis, a significant difference in overconfidence bias was observed in the analysis of Experimental Phase, $F(3, 126) = 5.36, MSE = .01, p = .004, {\eta^2}_p = .11$. A significant effect of Categorization Rule was also observed, $F(1, 42) = 10.88, MSE = .02, p = .002, {\eta^2}_p = .21$. These findings suggest that participants' general awareness of the category structure differed between rule-based and information-integration category structures (see Table 1) as well as over experimental blocks of trials. As is evidenced in the nearly additive pattern in Figure 2, participants exhibited greater overconfidence in the rule-based condition relative to the information-integration condition. This finding would be expected if the participants in the rule-based condition created a representation that was more accessible to subjective awareness while failing to account for negative response feedback or exception exemplars that is used by an procedural-learning system. The pattern evidenced in Figure 2 also suggests that participants' overconfidence bias generally decreased over experimental blocks. This pattern suggests that the rapid generation of a representation within the hypothesis-testing system might lead participants to believe that their performance was in fact better than it was in early phases of the experiment (i.e., they "knew" what the categories looked like). With additional trials, participants' accuracy eventually caught up to their subjective confidence, thereby reducing overconfidence.

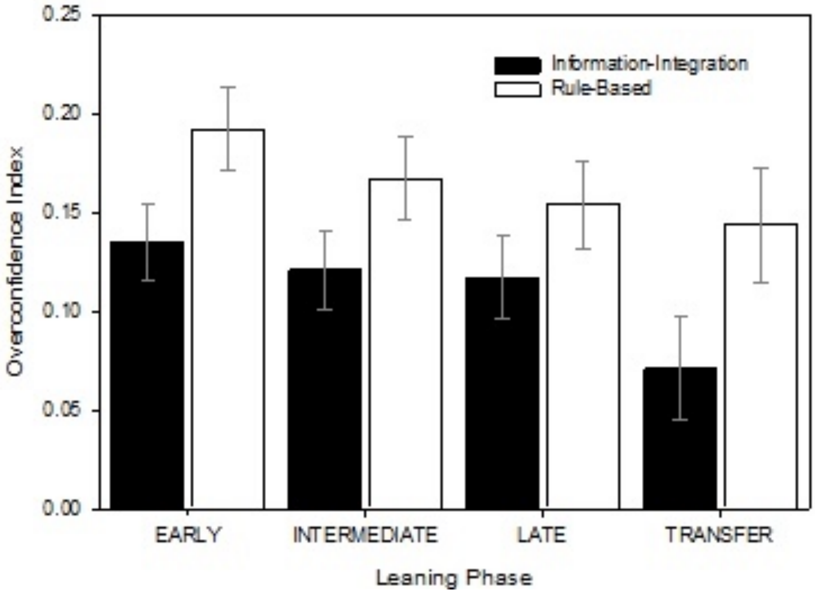


Fig. 2. Overconfidence bias for rule-based and information-integration category structures. Maximum overconfidence located at .35 (not plotted). The error bars represent the standard error of the mean. Phases reflected grouped sessions (e.g., Early: S1, S2; Transfer: S11, S12).

Conclusions

The results of our experiment replicated several earlier studies within the categorization and confidence processing literatures as well as our previous study (Schoenherr & Lacroix, 2012). Replicating Ell and Ashby (2006), while we found that learning was affected by categorization rule and expeirmental block, these two factors did not interact. Relative to our previous study (Schoenherr & Lacroix, 2012), we observed greater overconfidence across training and transfer. Thus, while the representation in the explicit learning system appears to be as accessible as in our previous studies thereby keeping confidence relative high, participants fail to account for their low performance and negative feedback. Thus, confidence reports appear to be determined by factors beyond those associated with the process of primary decision evidence accumulation and response accuracy. Taken together with our previous results, these findings suggest that multiple learning system acquire representations in a qualitatively different manner.

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. *Psycho-*

logical 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.

Ell, S. W.& Ashby, F. G. (2006). The effects of category overlap on information-integration and rule-based category learning. *Perception & Psychophysics,* 68, 1013–1026.

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.

Paul, E. J., Boomer, J., Smith, J. D., & Ashby, F. G. (2011). Information-integration category learning and the human uncertainty response. *Memory Cognition,* 39, 536–554.

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. *Attention, 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.

CAPTURING DECISION CONFIDENCE THROUGH RESPONSE TRAJECTORIES AND WILLINGNESS TO GAMBLE

Arkady Zgonnikov[1,2], Aisling Kenny[1], Denis O'Hora[1], KongFatt Wong-Lin[3]
[1]*School of Psychology, National University of Ireland Galway, Galway, Ireland*
[2]*University of Aizu, Aizuwakamatsu, Fukushima, Japan*
[3]*Intelligent Systems Research Centre, Ulster University, Derry/Londonderry, UK*
<arkady.zgonnikov@gmail.com, denis.ohora@nuigalway.ie>

Abstract

We aimed to investigate whether action dynamics could be employed as an objective measure of decision certainty and the relationship between certainty and confidence. Twenty-eight participants were required to view a random dot kinematogram display and report the dominant dot direction by moving the computer mouse. Directly following this, they were required to report the amount of points they were willing to bet that the answer they gave was the correct one. Coherence of the stimulus was experimentally manipulated and participants were required to complete 11 experimental blocks, each containing 48 trials of varying dot coherence. Mouse trajectory information was not predictive of post-decision certainty but was strongly related to decision accuracy. The findings were in line with a view of confidence as an evaluation of evidence which continues to accumulate after a decision.

Typically in our day to day lives, as we make a judgement or decision, it is followed by a subjective sense of certainty that the right decision was made. This feeling of certainty in our choices is commonly termed decision confidence. In the context of perceptual decision making, two current accounts suggest different mechanisms behind formation of confidence. Van den Berg et al. (2016) argue that confidence arises from the same evidence accumulation mechanism that underlies the formation of the original decision. In contrast, Murphy et al. (2015) propose that confidence is a product of a higher-order meta-cognitive process evaluating evidence beyond the initial decision. Most recently, Fleming and Daw (2017) proposed a model of second-order decision confidence, which generalizes and unifies these two approaches.

The current study investigated whether characteristics of participant's hand movements can act as an objective measure of on-line decision confidence. Comparing this measure to both performance accuracy and subjective confidence reports provides further insight into the process underlying retrospective confidence. Participants were required to make a perceptual discrimination task under differing levels of certainty and subsequently supply a measure of post hoc confidence. While previous research has often looked exclusively at associations between performance on a task and subsequent confidence judgements, the current study looks at associations between performance accuracy, confidence reports, and response trajectories. In this way, this research aims to provide another layer of evidence to the debates on the nature of decision confidence.

Methods

Undergraduate psychology students ($N = 28$, 4 male, 24 female, mean age 19.6 years) completed an experimental session in exchange for course credit. All study procedures

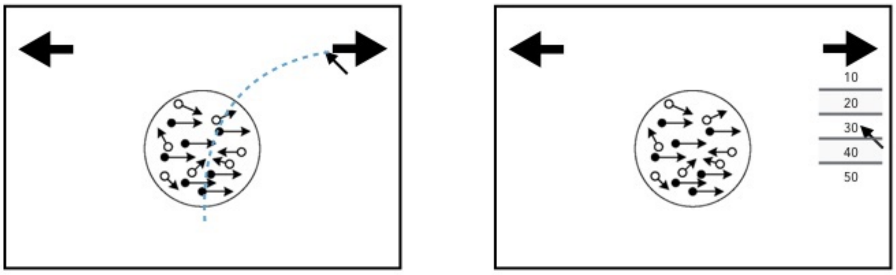


Fig. 1. Experimental setup: response screen during initial decision (left panel) and gamble selection (right panel)

employed were approved by the Research Ethics Committee at the National University of Ireland, Galway.

Participants made a series of 528 perceptual decisions, each followed by a "gamble". On each trial, participants were asked to judge the prevalent direction of a random dot kinematogram (RDK). The dot generation algorithm employed in this study was based on that of Shadlen and Newsome (2001) and programmed using PsychoPy (Peirce, 2007). The participants indicated their choice by moving the mouse cursor from the starting position at the bottom centre and clicking on one of the response locations in the top corners of the screen (Fig. 1). The RDK stimulus was present on the screen until one of the response locations was selected. Mouse coordinates during the response were recorded at 60Hz.

Motion coherence (the probability of any particular dot being displaced in the stimulus direction) constituted one independent variable and was manipulated within participants. The experimental session required participants to complete 11 trial blocks, each consisting of 48 trials. There were four coherence levels presented within each block (0.032, 0.064, 0.128, 0.256). All experimental blocks contained 12 trials of each coherence level, randomly shuffled, with stimulus direction (left or right) randomly determined for each trial.

On choosing a direction, participants were required to gamble 10 to 50 points on their answer using a drop-down menu. If the correct direction was chosen, a participant gained the chosen points, and if they chose the incorrect direction, the same number of points was deducted from their accumulated score. Participants were told at the beginning of the experiment that the aim was to earn as many points as possible by the end of the session. Upon making a number selection, a feedback stimulus appeared on the screen informing subjects whether they had answered correctly and the number of points won or lost.

Results and Discussion

The present analysis excludes the participants who could not perform the task accurately enough. Specifically, we excluded 12 participants who had accuracy below 75% at the coherence level of 0.256 and/or accuracy below 65% at the 0.128 coherence level.

The rest of the participants ($N = 16$), as expected, were more likely to correctly discriminate the direction of the RDK stimulus with increasing coherence (Fig. 2). In approximately 50% of the trials, the mouse trajectories indicated that the participants changed their preference during the course of the trial. Such trials were labelled as

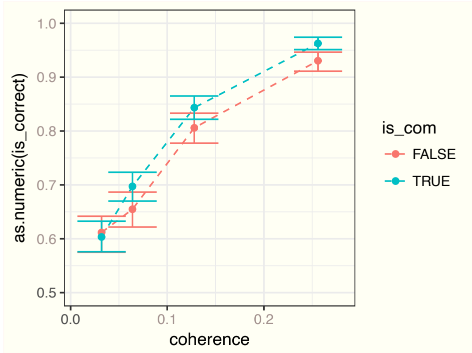


Fig. 2. Psychometric functions (averaged across 16 participants) for change-of-mind and non-change-of-mind trials.

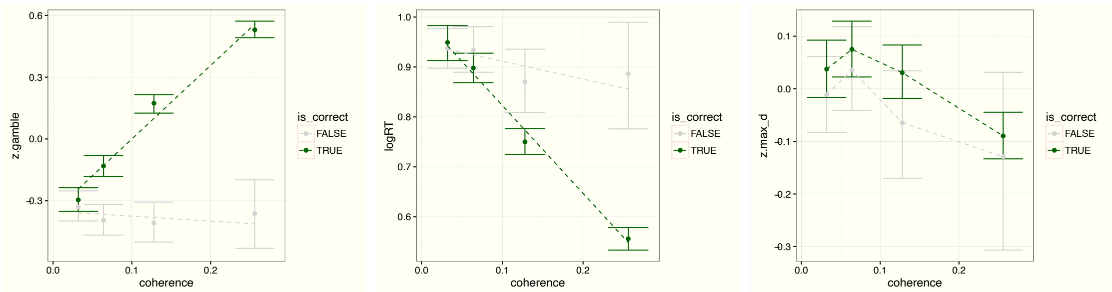


Fig. 3. Gamble value (z-scored), response time (log-scaled), and mouse trajectory curvature (z-scored) as a function of coherence for correct and incorrect trials. Correct responses are depicted in green and incorrect responses in grey. Whiskers denote bootstrapped 95% confidence intervals.

changes-of-mind. In accordance with the previous studies (e.g., Resulaj et al., 2009), changes-of-mind improved accuracy (Fig. 2). This might reflect the fact that after initial decision, additional evidence was continuously available to the participants until they clicked on one of the response locations.

The amount of points gambled after each decision increased with coherence for correct trials (Fig. 3, left panel). For incorrect trials, the amount gambled remained consistently small across coherence levels, which indicates that the participants could reliably detect their erroneous responses post-decision when the stimulus coherence was high (0.128 or 0.256). Together, these patterns suggest that gambled amount reflects subjective post-decision confidence of decision makers.

Previously in the mouse-tracking literature on value-based decision-making it has been suggested that mouse trajectories are linked to relative subjective value of the available options (e.g., McKinstry et al., 2008; Dshemuchadse et al., 2013; O'Hora et al., 2016). Here we hypothesize that in perceptual decision making, mouse trajectories may provide a measure of confidence within the response. To this end, we analyse response time and trajectory curvature as a function of RDK coherence.

As expected, response time decreased with coherence for correct responses (Fig. 3, centre panel). Error response time tended to remain high for all coherence levels. This is in line with consistently high values of gamble value, and thus reinforces the view of response times being related to decision confidence.

To measure trajectory curvature, we calculated maximum deviation (max-d) of each trajectory from the shortest trajectory towards the corresponding response area. In correct trials, max-d for correct trials exhibited non-monotonic relationship with coherence (Fig. 3, right panel). Although one might expect trajectory curvature to decrease as coherence increased (that is, as decisions became easier), mean max-d initially increased as coherence increased from 0.032 to 0.064. When coherence increased further to 0.128 and then 0.256, mean max-d decreased, in line with the expected pattern. As higher values of max-d in the present paradigm indicate greater rate of changes-of-mind, this finding is consistent with the post-decision evidence accumulation account of changes-of-mind (Resulaj et al., 2009).

Overall, our results suggest that changes of response direction during motor execution of a decision are informed by late-coming signal rather than by noise in the stimulus. Moreover, decision confidence as reflected in post-decisional wagering is related, but not equivalent to curvature of the response trajectories. Further investigations will shed light on the nature of relationship between within- and post-decision confidence.

Acknowledgements

This research was sponsored in part by Government of Ireland Postdoctoral Fellowship GOIPD/2015/481.

References

Dshemuchadse, M., Scherbaum, S. and Goschke, T. (2013), 'How decisions emerge: Action dynamics in intertemporal decision making.', *Journal of Experimental Psychology: General* **142**(1), 93.

Fleming, S. M. and Daw, N. D. (2017), 'Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation.', *Psychological Review* **124**(1), 91.

McKinstry, C., Dale, R. and Spivey, M. J. (2008), 'Action dynamics reveal parallel competition in decision making', *Psychological Science* **19**(1), 22–24.

Murphy, P. R., Robertson, I. H., Harty, S. and O'Connell, R. G. (2015), 'Neural evidence accumulation persists after choice to inform metacognitive judgments', *Elife* **4**, e11946.

O'Hora, D., Carey, R., Kervick, A., Crowley, D. and Dabrowski, M. (2016), 'Decisions in motion: Decision dynamics during intertemporal choice reflect subjective evaluation of delayed rewards', *Scientific Reports* **6**, 20740.

Peirce, J. W. (2007), 'PsychoPy—psychophysics software in Python', *Journal of Neuroscience Methods* **162**(1), 8–13.

Resulaj, A., Kiani, R., Wolpert, D. M. and Shadlen, M. N. (2009), 'Changes of mind in decision-making', *Nature* **461**(7261), 263–266.

Shadlen, M. N. and Newsome, W. T. (2001), 'Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey', *Journal of Neurophysiology* **86**(4), 1916–1936.

Van den Berg, R., Anandalingam, K., Zylberberg, A., Kiani, R., Shadlen, M. N. and Wolpert, D. M. (2016), 'A common mechanism underlies changes of mind about decisions and confidence', *Elife* **5**, e12192.

Part VIII
Free Talk Session 5

EFFECT OF SOUND ON MEMORY AND IMPRESSIVENESS OF VISUAL IMAGERY

Natalia Postnova and Shin-ichiro Iwamiya
Communication Design Science Course, Department of Design
Graduate School of Design, Kyushu University, 4-9-1 Shiobaru, Minami-ku
Fukuoka 815-8540, Japan
<nataliapostnova@gmail.com, iwamiya@design.kyushu-u.ac.jp>

Abstract

To clarify the functions of sounds in multimedia and the effect of sound on memory of visual imagery, the psychophysical experiments were conducted. The experimental stimuli were a series of symbolic pictures presented in video sequences with or without sound, or half with/half without sound. After every sequence participants were asked to recall the symbols they saw by selecting the right ones from the list. Experiments showed that the number of correct answers was larger when all the pictures in the sequence were presented with sound than when all or half of the pictures in the sequence were presented without sound. The same tendency was noticed when half of the symbols in the sequence were presented with sound and half without, the participants reported a higher number of symbols that were presented with sound. Furthermore, in the same condition, the number of selected pictures as the most impressive was larger when the pictures were presented with sound. These results suggested that sounds in multimedia have functions to emphasize the visual impression and facilitate visual memory.

The development of the contemporary media raised a lot of questions on the effect of sound on perceived imagery information. The memorability and impressiveness of the visual information is an important aspect of media-production. Can sound boost our memory or on the contrary distract us from the visual information? The increasing use of multimedia materials for various purposes including educational, therapeutic, recreational and others makes these questions even more important to answer.

Previous research in this area showed a complex relation between sound as a distracter and sound as a supporter of visual information. On one hand, we know that speech and non-speech stimuli of non-steady state can disrupt the performance of simple cognitive tasks (Jones et al., 1993), but on the other hand congruent sound can benefit the perception of visual information (Vroomen et al., 2000). However, the effect of the congruent, but non-steady sound is questionable. The purpose of the current study is to clarify the effect of a simple sound on memory and impression of visual imagery. To measure the expected effect, a simple memory span test is used (Miller et al., 1956; Ellermeier et al., 2015). We expect to detect difference in performance depending on the presentation of the visual imagery—with sound or without sound. As the presentation of the visual information in the memory span tests is usually periodical—the information appears on the screen one by one over equal periods of time, the sound synchronized with every element is steady periodical sound. To distinguish between the effect of the steady and non-steady sound, the third condition is used. In that condition, not all of the visual elements of memory span test are presented with the sound, but just a half of them. That way we can distinguish the effect of the sound itself, and the effect of steady and non-steady sounds on the performance of the cognitive task.

Fig. 1. Symbolic images used in the experiments.

Selection and designing of materials for visual stimuli

In order to conduct the experiments, the visual and sound materials for the memory span test had to be designed. As the test was based on the standard memory span test, the visual information for memorization should have been decided. Usually the digits are used—they are presented in a sequence of 3 to 10 digits once (depends on participant aptitude), and participants are asked to recall it in a correct order. During the pre-experimental stage, several stimuli were tested—digits, roman letters (consonants only) and black-and-white symbolic images. In the cases of digits and roman letters, participants were found to use various memorization techniques—for example, vocalization of the digits or letters and making meaning of the sequences. That consequently compromised the results. In the case of symbolic images, uses of such techniques were not found, hence the black-and-white symbolic images were chosen as visual stimuli for the experiments. The images are shown in Figure 1 (Fig. 1).

In the final design, 20 symbolic images were used to create 12 random sequences of 8 symbols each. For each sequence 3 sound conditions were created:

1. No sound (NS condition): control condition.

2. Fully synchronized sound condition (SS condition): Every symbol appearance was synchronized with the sound—440-Hz pure tone (0.1 second).

3. Half synchronized sound condition (HSS condition): Randomly chosen half of the symbols in the sequence were presented with sound (the same as in SS condition), while others without.

Figure 2 (Fig. 2) shows the schematic representation of SS and HSS conditions.

All together there were 36 videos—12 sequences of 8 symbols each in 3 different conditions. The same video material was used for 1st and 2nd experiments. Both experiments were conducted in soundproof rooms, and the sound stimuli were presented using headphones with the A-weighted sound pressure level of 67dB in both ears.

The experiments consisted of 3 sessions with 5 minutes breaks in between. Every session consisted of 12 different videos in different conditions, so during 3 sessions all 12 sequences were presented in all 3 conditions.

Experiment 1: The effect of sound on memorization of visual imagery

The purpose of the first experiment was to detect the effect of sound on short term memory. The participants were asked to watch the sequences and try to memorize the

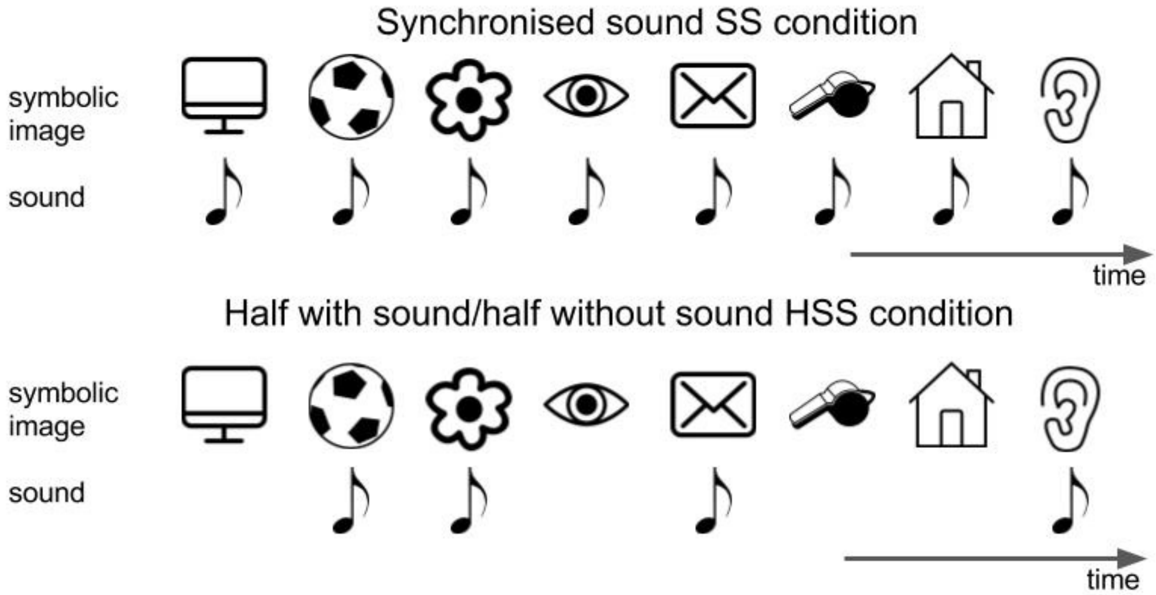


Fig. 2. Sound conditions in SS and HSS.

symbols (order was not important). After each sequence, they were asked to mark the symbols they remembered on the answer blank.

The scores were calculated based on the number of correct symbols recalled, the results were grouped based on the sound condition the sequence was presented in. To compare performances in different conditions, the absolute score for each participant was normalized against the controlled condition—NS (no sound) condition. For each participant, the result of performance in NS conditions in the session was calculated and performances in other conditions—SS and HSS, were measured against its score in %. This normalization was used in order to eliminate individual differences of performance. The average performance among all the participants in different groups using normalized scores are presented in a table (Table 1).

Because the 95% confidence interval of average value in SS condition does not include 100, the difference of performance between SS and NS condition is statistically significant. As it can be seen from table (Table 1), participants performed higher in SS condition. On the other hand, HSS condition did not have the same effect. The performance level in HSS is equivalent to that in NS condition. The difference of average

Table 1. The average performance in different conditions in percentages to NS condition (HSS—half synchronized sound condition, SS—fully synchronized sound condition, NS—no sound controlled condition)

Condition	Average (±95% confidence interval)
HSS	99 (±4)
SS	108 (±7)
NS	100

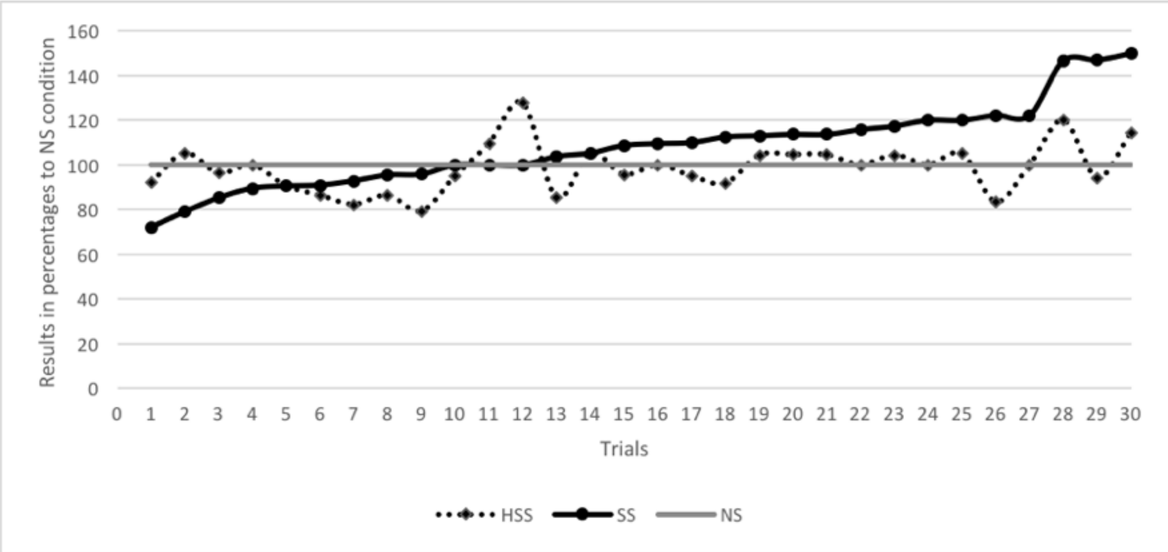


Fig. 3. Memory test performances results in each trial. HSS—half-synchronized condition, SS—fully synchronized condition.

values between SS and HSS condition is statistically significant, $t(47) = -2.4, p = 0.02$.

The comparison graph, showing the result of every trial in SS and HSS conditions compared with NS conditions is presented in Figure 3. As it can be seen from the graph, the SS condition on average has better results than performances in other conditions—in 23 out of 30 trials performance in SS was better than performance in HSS, and in 18 out of 30 trials performance in SS was better than in controlled NS condition.

In HSS condition the effect of sound on memory can be compared within the same trials, as half of the stimuli were presented with sound and half without. These two numbers—the number of recalled symbols that were presented with sound and the number of recalled symbols that were presented without the sound, were calculated separately. This analysis showed that the percentage of correctly recalled symbols with sound was higher than the percentage of symbols without sound. These numbers are presented in Figure 4.

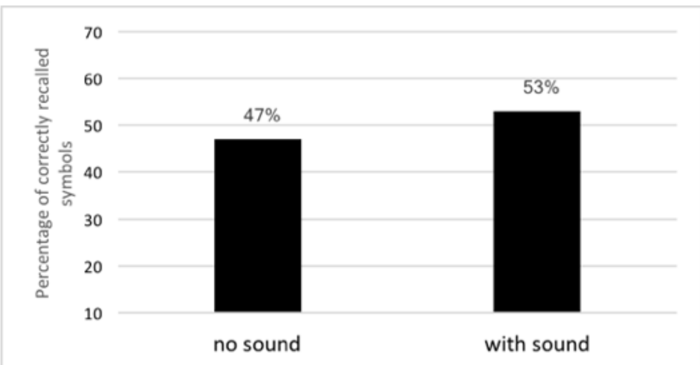


Fig. 4. Correctly recalled symbols with sound and without sound in HSS condition (% from overall number of the recalled symbols).

It is seen that participants reported on average 6% more symbols presented with sound than those presented without the sound, and the difference between them is statistically significant, $t(70) = 2.6, P = 0.011$.

Based on these results, we can conclude that sound affected the performance of the task of the memorization of visual stimuli. Results showed that sound in general made the visual stimuli more memorable. The steady fully synchronized congruent sounds boosted the performance of the task, while non-steady sound did not have such effect. However, the tendency to memorize stimuli with the sound better remained even in case of non-steady sounds.

Experiment 2: Effect of sound on the impressiveness of visual imagery

The purpose of the 2nd experiment was to clarify the effect of a sound on the impressiveness of the visual imagery. For this experiment, the same visual materials as in experiment 1 were used. The participants were asked to watch the video sequences without trying to memorize the symbols, and after watching draw 3 symbols that seemed to be most impressive of all of them.

The results were analyzed with the following method. Symbols in every video sequence were separated in 2 groups based on how particular symbols in particular sequences were presented in HSS (half synchronized) condition: S group (with sound) and N group (without sound). The presentation condition of symbols of S and N groups was different only in HSS condition; in SS condition both groups were presented with the sound and in NS condition both groups were presented without sound. So that way if there is no difference between the number of drawn symbols from S-group and N-group in conditions NS and SS, but there is in condition HSS, then we can say that it is the result of the effect of the sound.

After the experiments, all the data was analyzed using the following method. For every participant the number of drawn symbols from S-group and the number of symbols from N-group was calculated for every condition and converted into percentages. The average percentages for each group among the participants are presented in Figure 5.

The graph shows that among the drawn symbols in HSS condition there were 20% more symbols that were presented with sound (drawn symbols from S group) than the ones that were presented without sound (drawn symbols from N group), and this difference was statistically significant, $t(11) = 4.3, P < 0.01$. As the same sequences were presented in different sound conditions it would be logical to assume that if the symbols themselves were more impressive because of some other features rather than the sound, we would see the same tendency in other conditions. However, the graph shows that in condition SS and NS the proportion of the group of symbols that were presented with or without sound in condition SN are almost 50/50, and the differences between S and N groups were not statistically significant. Therefore, we can conclude that synchronized sound affect the impressiveness of the visual imagery.

Discussion

Altogether 30 trials were conducted for each of the experiments, and performances were analyzed using statistical analysis. The results showed that fully synchronized congruent sounds boosted the performance of simple memory task. In the follow up interview after the experiment, some participants reported that NS (no sound) condition was the most

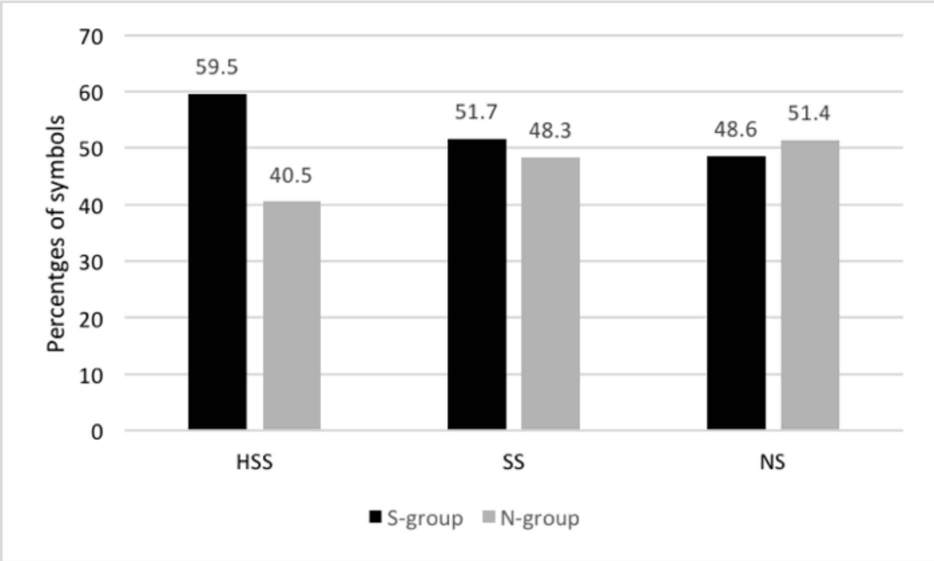


Fig. 5. The percentage of drawn symbols from S (with sound) or N (without sound) groups in three different conditions. HSS—half synchronized sound, SS—fully synchronized sound, NS—no sound.

complicated for them as it was difficult to concentrate, while others reported that HSS (partially synchronized) condition caused the most confusion and distraction, as non-steady sound was "annoying." No one reported SS condition as the most difficult one, though some participants mentioned that they preferred the condition without any sound.

Based on the results, it was suggested that we dealt with two different effects here. One is the effect of sound itself and another is the effect of periodicity of sound. It is known that non-periodic sounds are much more distractive than periodic ones, and non-periodic sounds attract our attention much more than periodic ones to the sound itself (Jones et al., 1993). We could see that effect when comparing half synchronized HSS condition and fully synchronized SS condition—participants performed worse in HSS than SS condition. It seems that periodic sounds boosted steady attention to the visual elements synchronized with it, because the performance in SS condition was higher than in no sound NS condition. Non-steady sounds in condition HSS also had the similar effect—we could see higher number of the symbols that were presented with sound reported, but as it was said earlier, in general non-periodic sound had more of a distracting effect and the performance was not improved by half synchronized sound compare to the NS condition. The further research can be conducted to confirm or reject these assumptions.

The second experiment showed the effect of sound on the impressiveness of symbolic images. The impression was stronger when images were presented with sound. This might facilitate the function of short term memory. The results of the 1st experiment support that: the number of recalled symbols presented with sound in HSS condition was larger than that without sound; the performance difference between SS and NS condition also showed the function of sound to facilitate the function of short term memory.

Another direction of further research could be focused on differences among participants, as some participants showed much higher influence to different sound conditions than others. It seems like some people are much more sensitive to sound, while others showed very small difference in performance. Further research can be done in order to clarify whether certain personal qualities, past experience or skills affect person's sensitivity to it.

References

Ellermeier, W., Kattner, F., Ueda, K., Doumoto, K. & Nakajima, Y. (2015). Memory disruption by irrelevant noise-vocoded speech: Effects of native language and the number of frequency bands. *The Journal of the Acoustical Society of America,* 138, 1561–1569.

Jones, D. M., Macken, W. J., & Murray, A. C. (1993). Disruption of visual short-term memory by changing-state auditory stimuli: The role of segmentation. *Memory and Cognition*, 21(3), 318–366.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97.

Vroomen, J., & de Gelder, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance*, 26(5), 1583–1590.

INVESTIGATING DIFFERENCES IN THE DYNAMIC-SYSTEMS STRUCTURE OF AUDITORY COGNITION AS A FUNCTION OF MUSIC TRAINING

Naomi du Bois[1], José M. Sanchez Bornot[2], KongFatt Wong-Lin[2], Mark A. Elliott[1]*, and Girijesh Prasad[2]
[1]*School of Psychology, National University of Ireland, Galway, Ireland*
[2]*Intelligent Systems Research Centre, University of Ulster, Magee Campus, Derry, NI*
*Correspondence: `<mark.elliott@nuigalway.ie>`

Abstract

A magnetoencephalographic (MEG) investigation of differences in the structure of the dynamic auditory cognition system, dependent on music training as opposed to no experience playing an instrument, was conducted using an auditory priming paradigm designed by Aksentijevic, Barber and Elliott [(2011). JEP: Human Percept. Perform., 37, 1628]. This paradigm employed stimulus entrainment to evoke an auditory gamma-band response (aGBR, i.e. an oscillatory response in the range 30-70 Hz) that is phase locked to the stimulus. Frequencies in this range have been demonstrated to facilitate a response to a deviant stimulus (inharmonic) depending on their relationship in phase with a slower theta rhythm [Elliott (2014). Front. Psychol.*, 5, 990]. Neuroscientific research has demonstrated that syntactically irregular chords elicit event related potentials (ERPs) with negative polarity and peak latencies of around 150-350 ms post stimulus onset [Rohrmeier & Koelsch (2012). Int. J. Psychophysiol., 83, 165-175]. The focus of the time frequency analyses for this was on the effect of priming on these auditory responses as a function of musical experience.*

Auditory binding refers to the integration of sounds to form a whole sound object from different acoustical components that are coded in anatomically separate parts of auditory cortex. This may be achieved through the synchronized activity of neural assemblies across the separate cortical loci with an emergent oscillatory code serving to bind neurons into a single dynamic assembly. Our research used an auditory priming paradigm developed by Aksentijevic, Barber, and Elliott (2011) to examine potential differences in the oscillatory codes involved in this process of binding, dependent on musical training. The paradigm primes the auditory system with sound stimuli designed to evoke an auditory gamma-band response (aGBR) that is phase locked to the stimulus. This is referred to as stimulus entrainment. Entrainment occurs when two independent oscillatory mechanisms become coupled due the alignment in phase of their respective periods, and it is an important characteristic of interacting brain rhythms. Gamma-band synchronization is considered to increases the strength of entrainment on target neural networks, establishing exclusive neural communication links necessary for cortical computation, while the influence of gamma-band synchronization can be either general or specific depending on how it is modulated (Fries, 2009). It has been proposed that the function of gamma band activity with regard to memory processes lies in the fine temporal precision required for the induction of neural coupling and long-term potentiation to create learning-dependent changes (Axmacher et al., 2006). Thus it is hypothesised that the paradigm used primes the system with an oscillatory code, reasoning that if it is an effective code responses to target sounds will be faster depending on the harmony relationship that is expected, and violations to these expectancies.

Music cognition concerns the cortical processes underlying our appreciation of music (Justus & Bharucha, 2002). According to Rohrmeier and Koelsch (2012) expectation and prediction (reliant upon temporally accurate protentive coding) are necessary for music perception. This research aims to apply previous findings regarding the temporal dynamics involved in cortical binding to an examination of the effects of priming on responses to violations of these expectancies. With regard to harmonic and melodic prediction, the EEG has revealed that syntactically irregular chords elicit ERP's with negative polarity and peak latencies of around 150-350 ms post stimulus onset (Rohrmeier & Koelsch, 2012). Two such predictive mechanisms are of interest to our proposed research; the early right anterior negativity (ERAN), and the right anterior temporal negativity (RATN) responses. Both are passively evoked responses to deviant stimuli elicited by music-syntax violations. However the RATN has only been observed in the cortical activity of musicians (Rohrmeier & Koelsch, 2012). Specifically our research hypothesised differences in the neural networks recruited by musicians and non-musicians during pitch processing, while revealing an increase in the strength of the RATN response in musicians due to musical training. In addition to the neural recruitment involved in this process, the behavioural evidence strongly suggests an interaction between the evoked aGBR's and a 6.69-Hz theta rhythm (possibly others) (Elliott, 2014). This research aims to establish convergence between the behavioural evidence and neural activity with regard to the involvement of neural networks in pitch processing and interactions in phase between emergent oscillatory systems and endogenous brain rhythms.

Method

Participants

Adult participants (music group (n = 6), ranging from beginner to Grade 8, and non-music group (n = 6)) were prepared for the MEG scanner.

Stimuli

A sequence of two sound stimuli (pip-trains) were used. The first was the entrainer, which carried a repeated sequence of a four pip prime; one 1,000 Hz deviant pip followed by three 500 Hz baseline pips. The second was the target stimulus which contained two different conditions—absent or present. The target absent (TA) pip-train consisted of only 1,000 Hz pips while the target present pip-train carried alternating pips which were either harmonically related to the entrainer (harmonic target present (HTP), 1000 Hz and 2000 Hz) or inharmonically related (inharmonic target present (ITP), 1000 Hz and 2400 Hz, in a ratio of 5:12). The diagram below (Fig. 1) illustrates the target conditions. Participants were instructed to respond as rapidly and accurately as possible to the presence or absence of the target tone in the second sound stimulus.

Design

The parameters of these stimuli were designed to replicate the stimuli in the paradigm which produced the previous 'pop-out' effect at 33 pips per second (pps) entrainment rate—that is slowing of reaction time (RT) responses to harmonic responses at a 33 pps

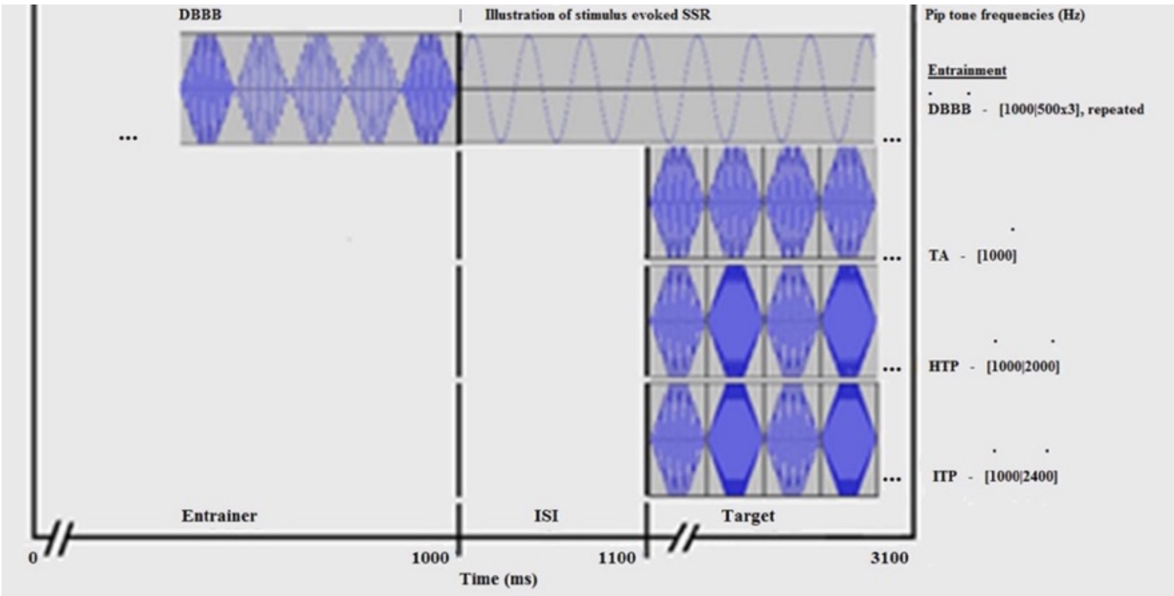


Fig. 1. Illustration of the target conditions for a 33 pips per second (pps) rate of entrainment. The pip train at the top of the diagram represents a repeated four pip entrainment sequence which is terminated at 1000 ms and comprises a total of 33 pips in this instance. The wave representation following this point is illustrative of the stimulus evoked steady state response (SSR). All three target examples are illustrated. Each is randomly selected during individual trials, and are presented following an interstimulus interval (ISI) of 100 ms.

entrainment rate, resulting in a deviation in the trend (see Aksentijevic et al. 2011, 2014, for details). This was a mixed design with a between group factor, and within group (5 x 2(2)) factors rate (31, 33, 35, 37, and 39 pps), and target (TA and target present—HTP and ITP).

Magnetoencephalography Measurement

Experimental sessions (1 per participant, $\sim$ 50 mins duration) were conducted at the Northern Ireland Brain Mapping (NIFBM) facility, at the Intelligent Systems Research Centre, Magee Campus, Ulster University, Derry, Northern Ireland. The continuous raw MEG was recorded per participant, per block (200 trials) using the 306-channel whole head MEG Elekta Neuromag system (Helsinki, Finland), comprising 204 gradiometers and 102 magnetometers. Ocular movements and cardiac activity were measured for cleaning purposes using four electrooculograph (EOG) electrodes (2 horizontal and 2 vertical), and one cardiac muscle electrode. Signals were digitised with a bandwidth of 0.1Hz to 300 Hz and a sampling rate of 1000 Hz. Sound stimuli were presented binaurally via ER-3A ABR Insert Earphones, and the decibel level was attenuated to 50 SPL, as measured by a Precision Gold (IEC 651 TYPE II) sound level meter (model #: N05CC).

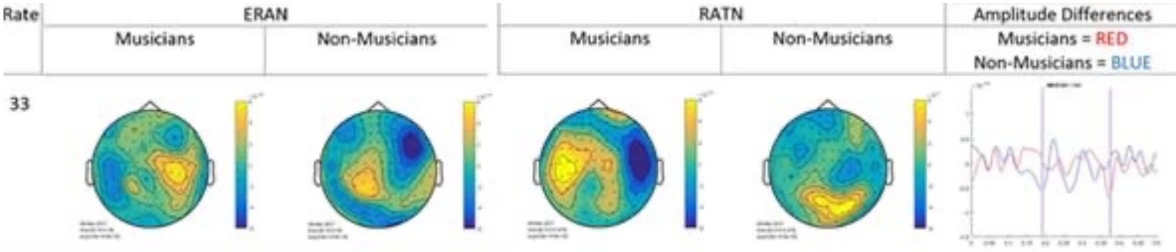


Fig. 2. Illustration of the 33 pps condition topography of the sensor level time frequency analysis for post stimulus latencies corresponding with the ERAN and RATN response windows (175-225 ms and 350-400 ms, respectively), for both groups, and the difference wave—at the MEG 1321 sensor. The difference wave (far right) is a measure of the difference in amplitude of cortical activity (event-related field's - ERF's) to harmonic and inharmonic stimuli from stimulus onset to 500 ms post-stimulus onset. The shaded line on the difference wave plots highlights the point in time corresponding to the topographic images preceding it.

Data Preprocessing

The continuous data were filtered off-line using Max Filter temporal signal space separation (tSSS). Data were cleaned and analysed using the FieldTrip toolbox in Matlab. Each participant's data were resampled to a rate of 500 samples per second (500 Hz), epoched, and concatenated across blocks. Cleaning included manual, squid-jump, and muscle artifact rejection, followed by removing outliers and running an independent component analysis (ICA) to remove the remaining components resulting from eye or cardiac movements. The clean data for all participants were averaged according to condition (HTP $\times$ Rate, and ITP $\times$ Rate).

Results and Discussion

Data Analysis at Sensor Level

Analysis of the difference waves for each entrainment rate by group condition, was conducted using sensor MEG1321, located over the right anterior temporal region. The RATN response was only observed in the cortical activity of musicians following an entrainment rate of 33 pps. The ERAN response appeared to be most clearly present among non-musicians at this rate also (Fig. 2).

A Mann-Whitney U test did not reveal a significant group difference for either response at any rate of entrainment, although the group difference at 33 pps for the RATN response was much closer to significance [p = 0.093, compared to other rates ($p \geq 0.3$)].Cluster analyses were conducted to determine significant event-related field (ERF) components during the latency windows corresponding to both the ERAN and RATN responses for each entrainment condition and group. A significant ERF was elicited within the RATN response window at 35 pps entrainment in the musician group (p = .02, latency of $\sim$ 365 ms), located in the right anterior temporal region, however the ERF component was positive. In order to examine whether there were differences in the RATN response as a result of experience playing a musical instrument, musicians data were

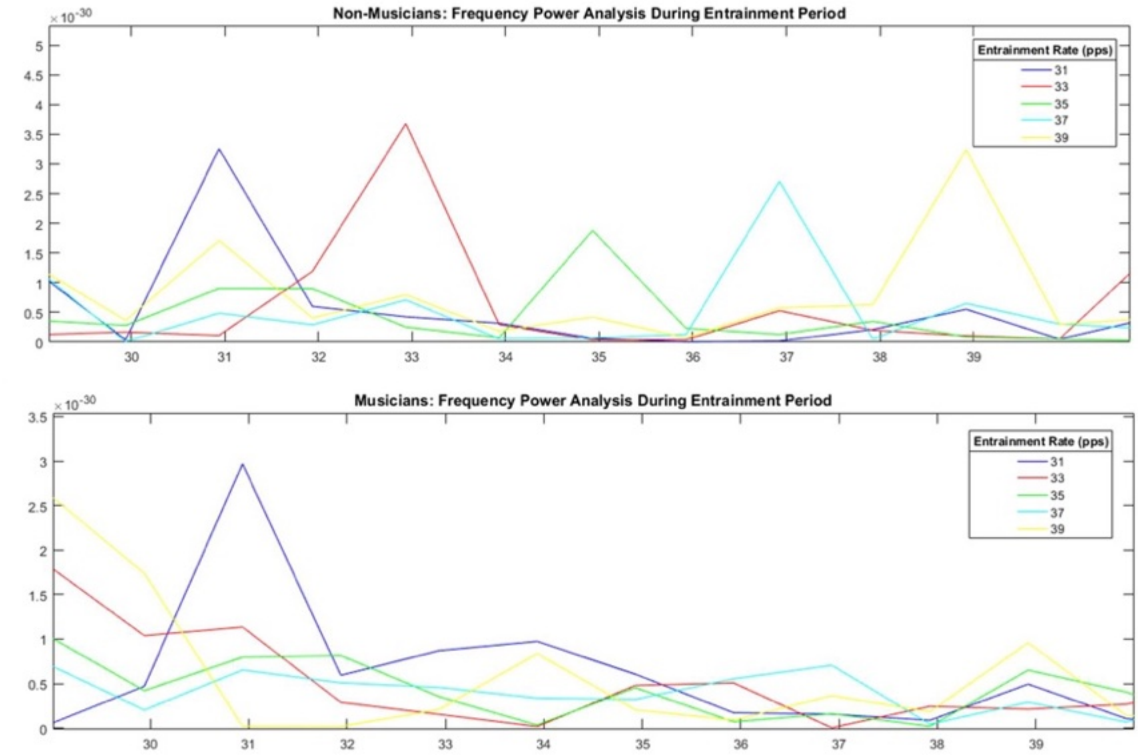


Fig. 3. Illustration of the amplitudes of frequencies between 20 and 50 Hz during the 100 ms ISI between entrainer and target, for each entrainment condition. Top: the graph for non-musicians. Bottom: the graph for musicians.

separated into two groups: grade 4 and lower ($n = 4$), and grade 8 ($n = 2$), and a cluster analysis was performed, again during the RATN response window. For both the lower grade and the grade 8 groups, at an entrainment rate of 35 pps, a positive ERF component was observed that was significant ($p = 0.0001$). For the lower grade group this is located in the anterior temporal to frontal region, with a latency window of between $\sim$ 360 ms, while for the grade 8 group this response occurs much more precisely at a latency of 366 ms and appears to be very strong, but is located in the right frontal region. For the grade 8 group only, positive ERF components, again significant ($p = 0.001$), were observed at entrainment rates of 37 and 39 pps, at later latencies of 398 and 388 ms respectively, located in the right posterior temporal, and right frontal regions respectively.

A frequency power analysis was conducted to examine group differences in the amplitude of the evoked response to the entrainer. A low pass filter of 50 Hz and a high pass filter of 20 Hz was used on the grand averaged data for each entrainment by group condition, with a window spanning the duration of the inter-stimulus interval (ISI) (-100 to 0 ms), covering right and left temporal MEG sensors. For the non-musician group clear peaks were revealed at the evoked frequency corresponding to entrainment rate. However for the musician group the only clear peak matching the corresponding entrainment condition appears at 31 Hz for the 31 pps condition. Peaks with a much lower amplitude ($\leq 1 \times 10 - 30$) were observed at corresponding frequencies for the 35, 37, and 39 pps entrainment conditions. However at the 33 pps rate of entrainment, there appears to be a complete absence of a 33 Hz signal (see Fig. 3).

The topography of the ERAN and RATN responses at the location defined in previous research (Maess et al., 2001), seemed to suggest that an entrainment rate of 33 pps (previously resulting in a pop-out effect) best facilitated an ERAN response in non-musicians and an RATN response in musicians (see Fig. 2). However group differences were not significant and the cluster analyses revealed that a significant response difference occurred during the 35 pps entrainment condition for the musicians only ($p = .02$), which was a right anterior temporal positive response. In support of this response, the behavioural data revealed that while musicians reaction time (RT) responses were faster in every condition compared to non-musicians, and responses to inharmonic targets were faster for both groups compared to harmonic responses—at 35 pps entrainment both groups' responded to both targets faster than at other rates, irrespective of harmony. Therefore at the sensor level, there are differences in the cortical responses to inharmonic stimuli dependent on musical training based on the observed significant positive ERF component, within the RATN latency window—however contrary to previous research this response is positive in polarity. Interestingly, this response became much stronger, more temporally specific (366 ms latency), and more frontal in location for the musicians with grade 8, which suggests that music experience does strengthen a response to deviant stimuli with a later latency than the ERAN response. According to Mussachia et al. (2007), musicians demonstrate practice related changes, i.e. larger and earlier (10 ms) post-stimulus onset ABRs, and enhanced phase locking to stimulus periodicity (presumed to underlie pitch perception), as well as enhanced representation of the fundamental frequency (f0). It has been suggested that this is due to a combination of top down influence and Hebbian learning. The former due to musical training modifying the neural architecture required for performance, beginning with higher processing areas and gradually enhancing lower sensory areas and the latter resulting from simultaneous activation of pre and post synaptic auditory brainstem neurons which strengthens the efficacy of the brainstem responses encoding sound. Thus music training leads to reciprocal efferent and afferent plasticity which strengthens the subcortical and cortical centres of the auditory corticofugal system.

The findings from previous research using a similar visual priming paradigm, thus far suggest a complex interplay between fast frequencies in the gamma range, and the modulating effects of slower rhythms (Elliott, 2014). Fast frequencies carry feedforward information such as stimulus input and prediction errors, and slower frequencies feedback templates and update the predictions (Michalareas et al., 2015). In this way anticipatory coding prepares the system to interact with the environment based on continuously updated object representations and predicted events. It may be that the entrainer benefits the non-musically trained brain by providing an anticipatory code in advance, while musicians are aided by well-established tone representations (templates), in higher cortical areas, and do not rely on the entrainer in the same way. The frequency power analysis provides some support for this theory, although the only entrained frequency observed was that corresponding to a 31 pps rate of entrainment, and this did not result in response facilitation. In addition the topography of the ERAN response in non-musicians and the RATN response in musicians appeared well defined at a 33 pps entrainment, while a 33 Hz frequency is almost absent during this entrainment condition. Aksentijevic et al. (2014) has argued that whereas rate-specific entrainment prepares non-musicians for target processing, in the case of a musically trained brain the entrainment frequency modulates the envelope of the carrier frequencies—the curve outlining the extremes of the oscillating signal, and at critical binding frequencies this facilitates a response overall. Given the findings of this research perhaps frequencies which are not critical for binding can disrupt the normal binding process in the musically experienced brain.

Further analyses are required, which will include a source analysis, using forward modelling, to look at differences in the strength of ERAN and RATN responses depending on group and entrainment frequency, and a cross frequency analysis to determine if there is gamma-theta coupling (during entrainment, and following target presentation). This latter analysis will address the question regarding an interaction in phase between gamma and theta, suggested by previous research, and may also shed more light on the results from the frequency power analysis.

Acknowledgements

This research was sponsored by the Irish Research Council (IRC), and the Northern Ireland Functional Brain Mapping (NIFBM) facility, Ulster University, Magee Campus, Derry.

References

Aksentijevic, A., Barber, P. J. & Elliott, M. A. 2011. Process timing and its relation to the coding of tonal harmony. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 1628.

Aksentijevic, A., Smith, A. & Elliott, M. A. 2014. Rate-specific entrainment of harmonic pitch - Effects of music training. *Music Perception: An Interdisciplinary Journal*, 31, 316-322.

Axmacher, N., Mormann, F., Fernandez, G., Elger, C. E. & Fell, J. 2006. Memory formation by neuronal synchronization. *Brain Res Rev*, 52, 170-82.

Elliott, M. A. 2014. Atemporal equilibria: pro- and retroactive coding in the dynamics of cognitive microstructures. *Front Psychol*, 5, 990.

Fries, P. 2009. Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annu Rev Neurosci,* 32, 209-24.

Justus, T. C., & Bharucha, J. J. (2002). Music perception and cognition. In S. Yantis (Volume Ed.) and H. Pashler (Series Ed.), *Stevens' Handbook of Psychology* (Third Edition, pp. 453-492). New York: Wiley.

Rohrmeier, M. A. & Koelsch, S. 2012. Predictive information processing in music cognition. A critical review. *Int J Psychophysiol*, 83, 164-175.

Michalareas, G., Vezoli, J., Van Pelt, S., Schoffelen, J.-M., Kennedy, H. & Fries, P. 2015. Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron*, 89, 384-397.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences*, 104(40), 15894-15898.

SOME TEMPORAL CHARACTERISTICS OF MUSIC COGNITION

Mark A. Elliott, Deirdre Farrell, Caitríona Eames, and Naomi du Bois
School of Psychology, National University of Ireland, Galway, Ireland

Abstract

We investigated whether pitch relations (chords) are processed via precise temporal mechanisms, and whether the use of temporal mechanisms depends upon musical training. Musicians and non-musicians were primed with an external gamma-band entrainer, presented at rates 31, 33, 35, 37 pips-per-second (PPS). Chord sequences were concurrently presented and included target chords (either consonant or dissonant) presented in or out of phase with the ongoing entrainer. Participants were asked to respond as rapidly as possible with their judgment of the consonance or dissonance of the target chord. Musicians responded significantly faster to out-of-phase dissonant targets and non-musicians responded significantly faster to out-of-phase consonant targets. A priming bias for out of phase targets suggests that cortical mechanisms coding harmony are precisely phase-locked to stimulus onset and thus analyses of harmony occurs at a very early stage of information processing. Oscillatory gamma activity presents itself as a candidate neural mechanism for the early processing of tonal features and the prospective coding of future tonal events.

The significance of synchronized gamma activity in auditory perception was brought to attention by Galambos, Maekig and Talmaschoff (1981), who reported auditory steady state responses (ASSR) to be particularly prominent at 40-Hz (mid gamma range). This has since been demonstrated with a sequence of clicks (Tiihonen, Hari, Kaukoranta, & Kajola, 1989), tone-pips (Gutschalk, Oldermann, & Rupp, 2009), and amplitude-modulated tones (Picton, Skinner, Campagne, Kellet, & Maiste,1987). In a recent study, Aksentijevic, Barber, and Elliott (2011) found that external oscillation entrainment in the gamma band at 33 pips per second (pps ie at 33 Hz) expedited reaction times (RTs) to harmonic targets carrying fractional multiples of the priming frequency. The induced gamma activity caused an "inharmonic popout" or increased salience of harmonic partials that were that were fractional multiples of the primed tones. Aksentijevic et al. (2011) proposed a model of harmony processing in which incoming resolved spectra are compared against internalized harmonic templates (formed via oscillatory synchronization). We assume this model to generalize to pitch processing, where pitch is determined according to the best matching templates.

Once pitch is coded, it is perceptually organized along a horizontal (e.g. melody) and vertical (eg. harmony) axis. In our musical system, pitch combinations (intervals and chords) exist in a hierarchy that determines their music-structural importance and relevance in musical piece (Lerdahl & Jackendoff, 1983). The rules of tonality specify the arrangement of chord functions within harmonic progressions (Krumhansl and Toiviainen, 2001). Chords that obey the rules of tonal harmony are regarded as consonant (subjectively pleasant or stable sounding) while those that violate the rules are regarded as dissonant (subjectively harsh or requiring resolution). Most of the music composed and produced today presupposes a sensitivity to tonal music syntax and despite an inability to articulate musical intuitions, untrained listeners are nevertheless capable of making judgments about chords and their relations that are highly sophisticated and consistent with music-theoretic descriptions (Unyk & Carlson; 1987, Tillman & Lebrun-Guillard, 2006). Recent neuroimaging studies, using harmonic violations to investigate tonal processing, have suggested that the extraction of harmony involves the activity of multiple brain areas. While it is unclear whether these regions operate in isolation or interact, it is speculated that communication between these brain areas would be characterized by transient phase relationships between oscillatory activities of underlying neuronal populations (Ruiz, Koelsch & Bhattacharya, 2008).

In the current study, we hypothesized that external gamma band entrainment (see Aksentijevic et al., 2011) presented concurrently with a sequence of chords would facilitate the communication between brain areas responsible for harmony processing. We hypothesized that the neuronal mechanism involved in the processing of tonal harmony occurs at an early stage of information processing and is therefore phase-locked to the stimulus. On this basis we hypothesized that we could open/close temporal windows for communication by presenting a target in phase or out of phase alignment with the gamma prime. We expected that the presentation of targets in phase with gamma entrainment would facilitate the exchange of information between oscillating neural assemblies and that the presentation of targets out of phase with the entrainment would inhibit the exchange of information. Based on previous findings of prospective coding mechanisms (Aksentijevic et al., 2011; Elliott, 2014) and the significance of 33 Hz in Aksentijevic and Aksentijevic-like paradigms we hypothesized that maximal priming potential would occur at 33 Hz. In studies of pitch extraction, it has been observed that gamma priming is not found in musicians (Aksentijevic, Northeast, Canty & Elliott, 2013). Given that the perceptual nuances of tonal harmony are augmented with experience (McDermott, Keebler, Micheyl, & Oxenham, 2010, Bidelman, Gandour & Krishnan, 2011), we used a broad sample of musicians and non-musicians to examine the interaction between stimulus frequency (31–37 Hz) and perceptual harmony. We expected that musicians would have a reduced reliance on external gamma entrainment.

Method

Participants

Twenty-eight participants took part in this study (12 female, mean age = 21.52 years, SD = 2.3 years); 14 were musicians, 12 of whom had received 2 to 14 years of formal training (mean = 6 years, 4 months), while 2 musicians were self-taught over a 10–14 year period. All musicians were trained on pitched instruments. Five participants were classically trained. Fourteen participants were non-musicians and reported no engagement with music beyond listening. All participants reported normal hearing and provided informed consent. The study was approved by the Ethics Committee of the Department of Psychology, National University of Ireland, Galway.

Apparatus and Stimuli

Experimental trial and stimulus generation were controlled by an IBM compatible computer. Trial presentation and data collection were implemented using Superlab (v.4, Cedrus). The stimuli were presented diotically via Beyerdynamic DT 880 Pro headphones. Participants responded to experimental stimuli on a Cedrus RB x360 response pad. Each trial consisted of a tone-pip train (prime) and overlapping chord sequence. The tone pip train was designed to generate an evoked gamma-band oscillatory steady state response

(SSR; Galambos, Makeig & Talmachoff, 1981) and was delivered in 4 frequency conditions [31, 33, 35 and 37 pps (pips per second)]. Pip duration was a reciprocal of presentation rate (e.g. 30.3 ms at 33 pps). Individual tone pips were generated in Audacity 2.0.5 and were offset and onset-ramped with symmetrical ramps that had a plateau of 33% of the overall period. The pip trains used in this study were modeled on the entrainment described in Aksentijevic et al., (2011).

A tone-pip train was presented independently at the beginning of each trial. Following 17–19 tone pips (approx. 515 ms), a sequence of chords was introduced and presented over the remaining pip train. The timing of sequence onset (17–19 pips) varied between rate conditions to ensure that the initial chord was optimally aligned with the phase of the entrainer. All entrainers and chord sequences carried 500 Hz and were presented at the intensities of 61 dB and 64 dB, respectively. The adjustment of intensity was carried out to control for the potentially distracting nature of the entrainer on task performance. Two chord sequence-types were presented: chord progressions that concluded on a consonant chord and progressions that concluded on a dissonant. The task was presented as a discrimination exercise for which participants were asked to respond to consonant and dissonant chord closures. Both stimulus-types consisted of the same chord progression and differed only with respect to their terminating chords.

Sequences varied in length from four to five chords. Consequently, there were four categories of chord sequences: a short sequence with a consonant closure, a longer sequence with a consonant closure, a short sequence with a dissonant closure, a longer sequence with a dissonant closure. This experimental manipulation was carried out to offset the repetitive nature of the task thus controlling for possible incidental entrainment of brain activity related to attentional deployment, in the EEG alpha band (i.e. 8–12 Hz). Both targets (consonant and dissonant closures) were presented an equal number of times. Consonant terminal chords were Tonic chords (I). The harmonic progression in the consonant condition was dominant to tonic (V–I). This is known as a perfect cadence and is the most harmonically appropriate progression at the end of a sequence. A perfect cadence commonly serves as the final event in a musical piece in that it elicits a feeling of closure and completion.

Dissonant terminal chords were Neapolitan sixth (N^6) chords. Due to its unusual half step root relationship to the tonic and overall distant relationship to the established musical key, the Neapolitan is perceived as striking and dissonant. The harmonic progression in the dissonant progression was dominant to Neapolitan (V–N^6). This is a highly inappropriate progression as the presentation of the dominant sets up the expectancy for a tonic. This progression therefore elicits a high degree of dissonance and surprise. Individual chords were created and exported as WAV files using composition and notation software Musescore 2. All chords were set to acoustic piano timbre to mimic an authentic musical stimulus. The chords (WAV form) were imported into Audacity 2.0.5 and inaudible fade-in/fade-out signals were trimmed from the audio. Individual chords were superimposed onto pip trains (prime) in accordance with the phase alignment of tone pips. Each chord initiated at the onset of a tone pip and terminated at the offset of another tone pip).

Chords lasted roughly 250 ms, however, they were varied in length (i.e. tone pip count) to control for the entrainment of an extraneous rhythm induced by the temporal regularity of chord onset. In all trials, the terminal chord (target) was longer and was ramped at offset to allow for the sound to resonate (aiding recognition of target) and to prevent a constrictive release. Tone pips discontinued midway through the terminal chord, again to prevent the plosive effect of sudden pip release. Target (terminal) chords were placed precisely out of phase alignment with the phase of tone-pip presentations on 50% of trials and were placed in-phase on the other 50% of trials. This was intended to offset the gamma-band oscillatory steady state response (SSR) achieved by the entrainment.

Design and Procedure

The study employed a within-subjects design with targets (consonant closure, dissonant closure) and factors rate (31, 33, 35, 37 pps) and alignment (in phase/out of phase), with the between-subjects factors musician and non-musician. There were 40 trials per condition resulting in 640 trials per participant. Trials were presented in eight blocks of 80 trials per block. The experiment was conducted in a sound-attenuated experimental cabin in a laboratory suite housed in the School of Psychology at NUI Galway. Stimulus intensity was held constant throughout [average 60 dB sound pressure level (SPL)], determined using an Adastraanalog sound level meter. All stimulus sequences were presented pseudo-randomly on a session wise basis. Each trial commenced with a blank screen. A randomized interval between two adjacent trials was set between 400 and 500 ms and was followed by stimulus presentation. Response time was measured from target offset and a new trial was initiated automatically from the response.

Magnetoencephalography Measurement

Experimental sessions (1 per participant, $\sim$ 50 mins duration) were conducted at the Northern Ireland Brain Mapping (NIFBM) facility, at the Intelligent Systems Research Centre, Magee Campus, Ulster University, Derry, Northern Ireland. The continuous raw MEG was recorded per participant, per block (200 trials) using the 306-channel whole head MEG Elekta Neuromag system (Helsinki, Finland), comprising 204 gradiometers and 102 magnetometers. Ocular movements and cardiac activity were measured for cleaning purposes using four electrooculograph (EOG) electrodes (2 horizontal and 2 vertical), and one cardiac muscle electrode. Signals were digitised with a bandwidth of 0.1Hz to 300 Hz and a sampling rate of 1000 Hz. Sound stimuli were presented binaurally via ER-3A ABR Insert Earphones, and the decibel level was attenuated to 50 SPL, as measured by a Precision Gold (IEC 651 TYPE II) sound level meter (model #: N05CC).

Results

Participants performed the task with high accuracy (96% correct for consonant and dissonant targets). Trials with error responses were removed from the data prior to analysis. Error RTs were slower than correct RTs and an analysis of the probability correct by RT showed no significant correlation between RT and accuracy. Examination of the correct RTs revealed non-normal distribution with a pronounced positive skew. RT distributions were approximately lognormal on a Kolmogorov 'D' test and as a result, subsequent analyses were performed on the exponents of the means of log-transformed RT distributions.

The RT means were subject to a three-way repeated measures analysis of variance (rANOVA) with the within-subjects factors rate (31, 33, 35, 37 pps), synchrony (in-/out-of phase) and harmony (consonant, dissonant) and the between-subjects factor musicianship (musicians, non-musicians). Violations of the homogeneity of variance assumption were corrected by applying Greenhouse Geisser or Huynh-Feldt adjustment criteria. There was a significant main effect for harmony $[F(1, 26) = 10.50, MSE = 5906, p < .005, {\eta^2}_p = .29]$ with consonant targets being registered more rapidly than dissonant targets [mean difference= 23.5 ms, standard error of the mean (SEM)= 7.26 ms]. The main effect of synchrony was significant $[F(2, 26) = 7.11, MSE = 5906, p < .05, {\eta^2}_p = .26]$ with faster RTs to targets that were presented out of phase alignment with the entrainer relative to those presented in phase (mean difference = 7.18 ms, SEM= 2.69 ms). There was a main effect for rate $[F(3, 78) = 6.48, MSE = 832, p = .001, {\eta^2}_p = .20]$, reflecting a tendency for RTs to be significantly slower when primed at 33 PPS. The effect of rate was influenced by synchrony as indicated by a significant rate by synchrony interaction $[F(3, 74.63) = 8.71, MSE = 731, p < .001, {\eta^2}_p = .25]$, resulting from a significant difference between RTs to targets presented in/out of phase alignment at 35 PPS (Mean difference = 13 ms, SEM = 2.65, $p < .001$) and 37 PPS (Mean difference = 23 ms, SEM = 4.64, $p < .001$) and other rates, 31 PPS (Mean difference = 9 ms, SEM = 6.55, $p > .5$) and 33 PPS (Mean difference = 9 ms, SEM = 5.70, $p > .5$). There was also a significant interaction between rate and harmony $[F(3, 76.83) = 3.7, MSE = 703, p < .001, {\eta^2}_p = .11]$, resulting from a significant difference in RT when presented with consonant/dissonant targets at 31 PPS (Mean difference = 34 ms, SEM= 4.64, $p < .005$), 35 PPS (Mean difference = 27 ms, SEM= 6.57, $p < .001$) and 37 PPS (Mean difference= 22 ms, SEM= 6.80, $p < .005$).

There was a significant three-way interaction between synchrony, harmony and musicianship $[F(1, 26) = 4.23, MSE = 437, p = .05, {\eta^2}_p = .14]$. Post-hoc analysis was carried out using a rANOVA with factors harmony, synchrony (collapsed over frequency) and musicianship. This analyses revealed the significant interaction to result from the combination of a significant RT advantage for musicians when presented with dissonant targets out-of-phase with the onset of entrainers (Mean difference= 8.76 ms, SEM= 3.40, $p < .05$) and an RT advantage for non-musicians when presented with consonant targets, out-of-phase with the onset of entrainers (Mean difference= 13.73 ms, SEM= 5.75, $p < .05$). Overall, there was no main effect, nor were there any other interactions of any other factors with musicianship, This indicates that musicians did not process any other stimulus characteristics differently to the non-musicians.

Discussion

Data Analysis at Sensor Level

A principle finding was of a general RT facilitation to targets that were presented out of phase with the entrainer. This effect could not be mediated strategically as participants were unable to report the phase synchrony of the prime. It suggests that neuronal temporal coding at gamma frequencies that are phase-locked to stimulus activity, mediates the extraction of perceptual harmony. We suggest that synchronized gamma oscillations activate the cortical neurons coding a given pitch and that these oscillations are extended to neuronal populations of related pitches. By extension, synchronized gamma activity may present a candidate neural mechanism for the generation of the total tonal experience (through phasic stimulation from pitch-chord-key units), as described in connectionist models of tonality (see Bharucha, 1987).

Contrary to hypothesis, a RT advantage to out-of-phase targets indicates that maximal priming was not achieved when targets were optimally aligned. Thus this condition did not facilitate the exchange of information between oscillating neural assemblies as expected. A general RT facilitation was observed for consonant targets. This could be attributable to the gamma entrainment or to the tonal context established by the harmonic progression (see Bharucha & Stoeckig, 1987). However, a baseline investigation of gamma entrainment vs. no entrainment may be necessary to verify this.

While there was no difference in overall RTs between non-musicians and musicians, musicians detected out-of-phase targets more efficiently when they were dissonant, non-musicians detected out-of-phase targets more efficiently when they were consonant. This indicates that the temporal dynamics of harmony processing differs between the two groups and supports recent evidence of differential gamma power (indicative of the number of synchronized neurons) in response to musical stimuli in musicians as compared with non-musicians (Bhattacharya, & Petsche 2001; Roberts, 2007). Given evidence of more robust and synchronous gamma activity in musicians, we expected that musicians would have an decreased sensitivity to entrainment at gamma-band frequencies and that consequentially, no effect would be observed. Contrary to this hypothesis, musicians RTs were sensitive to half phase manipulations in the prime, indicating that the entrainment did modulate the synchrony of ongoing cortical oscillations in musicians. The RT advantage for out-of-phase dissonant targets in musicians could be due to the function both of gamma oscillations as an anticipatory mechanism (Elliott, 2014), and the schematic discrepancy of the dissonant chord. If enhanced gamma activity reflects both the expectation and realization of predicted events, musicians may be more sensitive to the perceptual nuances of musical syntax and therefore experience a higher degree of schematic discrepancy when a syntactically unexpected event occurs. The harmonic progression of a dominant chord function to a Neapolitan chord (V–N^6) is extremely rare in modern composition and is therefore highly inconsistent with a (western) musician's schema of tonal relations.

References

Aksentijevic, A., Barber, P. J., & Elliott, M. A. (2011). Process timing and its relation to the coding of tonal harmony. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 1628–1642.

Aksentijevic, A., Northeast, A., Canty, D., & Elliott, M. A. (2013). The oscillatory entrainment of virtual pitch perception. *Frontiers in Psychology*, 4.

Bharucha, J. J (1987). MUSACT: A Connectionist Model of Musical Harmony. *Proceedings of the Ninth Annual Conference of the Cognitive Science Society*. Hillsdale,

Bharucha, J. J., & Stoeckig, K. (1987). Priming of chords: Spreading activation or overlapping frequency spectra?. *Perception & Psychophysics*, 41(6), 519–524.

Bhattacharya, J., & Petsche, H. (2001). Musicians and the gamma band: a secret affair? *NeuroReport*, 12(2), 371–374.

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain and Cognition*, 77(1), 1–10.

Elliott, M.A. (2014). Atemporal equilibria: pro- and retroactive coding in the dynamics of cognitive microstructures. *Frontiers in Psychology*, 5:990. doi:10.3389/fpsyg.2014.00990

Galambos, R., Makeig, S., & Talmachoff, P. J. (1981). A 40-Hz auditory potential recorded from the human scalp. *Proceedings of the National Academy of Sciences*, USA, 78, 2643–2647.

Gutschalk A, Oldermann K, Rupp A (2009) Rate perception and the auditory 40-Hz steady-state fields evoked by two-tone sequences. *Hear. Res.*, 257:83–92.

Krumhansl, C. L., & Toiviainen, P. (2001). Tonal cognition. *Annals of the New York*

Academy of Sciences, 930(1), 77–91.

Lerdahl, F., & Jackendoff, R. (1983). An overview of hierarchical structure in music. *Music Perception: An Interdisciplinary Journal*, 1(2), 229–252.

McDermott, J. H., Keebler, M. V., Micheyl, C., & Oxenham, A. J. (2010). Musical intervals and relative pitch: Frequency resolution, not interval resolution, is special. *The Journal of the Acoustical Society of America,*128(4), 1943–1951.

Picton, T. W., Skinner, C. R., Campagne, S. C., Kellet, A. J. C., & Maiste, A. C. (1987). Potentials evoked by the sinusoidal modulation of the amplitude and frequency of a tone. *Journal of the Acoustical Society of America* 82, 175–178.

Ruiz, M. H., Koelsch, S., & Bhattacharya, J. (2009). Decrease in early right alpha band phase synchronization and late gamma band oscillations in processing syntax in music. *Human brain mapping*, 30(4), 1207–1225.

Tiihonen, J., Hari, R., Kaukoranta, E., & Kajola, M. (1989). Interaural interaction in the human auditory cortex. *Audiology*, 28, 37–48.

Tillman, B. & and Lebrun-Guillard, G. (2006). Influence of tonal and temporal expectations on chord processing and on completion judgements of chord sequences. *Psychological Research*, 70, 345–358.

Unyk, A. J., & J. C., Carlson (1987). The influence of expectancy on melodic perception. *Psychomusicology*, 7, 3–23

Part IX

Poster Session 1

SCALING DEPRESSION WITH PSYCHOPHYSICAL SCALING: A COMPARISON BETWEEN THE BORG CR SCALE® (CR100, CENTIMAX®) AND PHQ-9 ON A NON-CLINICAL SAMPLE

Elisabet Borg and Jessica Sundell
Department of Psychology, Stockholm University, SE-10691 Stockholm, Sweden
<eb@psychology.su.se>

Abstract

A non-clinical sample ($n = 71$) answered an online survey containing the Patient Health Questionnaire-9 (PHQ-9), that rates the frequency of symptoms of depression (DSM-IV). The questions were also adapted for the Borg CR Scale® (CR100, centiMax®) (0–100), a general intensity scale with verbal anchors from a minimal to a maximal intensity placed in agreement with a numerical scale to give ratio data). The cut-off score of PHQ-9≥ 10 was found to correspond to $\geq 29cM$. Cronbach's alpha for both scales was high (0.87) and the correlation between the scales was $r = 0.78$ ($r_s = 0.69$). Despite restriction of range, the cM- scale works well for scaling depression with added possibilities for valuable data analysis.

According to the World Health Organisation (2016), depression is a growing health problem around the globe. Since it is one of the most prevalent emotional disorders, can cause considerable impairment in an individual's capacity to deal with his or her regular duties, and at its worst is a major risk factor for suicide attempt and suicide, correct and early diagnosis and treatment is very important (e.g., APA, 2013, Ferrari *et al.*, 2013; Richards, 2011).

Major depressive disorder (MDD) is characterized by distinct episodes of at least two weeks duration involving changes in affects, in cognition and neurovegetative functions, and described as being characterized by sadness, loss of pleasure or interest, fatigue, low self- esteem and self-blame, disturbed appetite and sleep, in addition to poor concentration (DSM- V, APA, 2013; WHO, 2016).

In parallel to the Diagnostic and Statistical Manual of Mental Disorders (DSM-V, APA, 2013) several instruments have been developed for assessment of depression and its symptoms, for example: the Hospital Anxiety and Depression Scale (HAD, Zigmond & Snaith, 1983); the Beck Depression Inventory (BDI, Beck *et al.*, 1961); the Montgomery—Åsberg Depression Rating Scale (Montgomery & Åsberg, 1979) and the Patient Health Questionnaire (PHQ-9, Kroenke, Spitzer & Williams, 2001). Weather constructed according to classical test theory or item response theory, the focus during construction of such psychological tests as these has been on the formulation and evaluation of test items, and not so much on the response scale used, commonly Likert scales of intensity, agree–not agree, or frequency. The test result for each individual is usually summed into a total score and compared to norms for the total population. Often, however, obtained data is of an ordinal character, more or less depression, and there is no saying how much more or less.

Within the psychophysical tradition of scale development, the main interest has been scales to obtain response data that grow linearly with perception (and sensation) when studying psycho-physical relationships (of, for example, loudness, brightness, taste, pain, exertion, etc.). One important method is Magnitude Estimation (ME) where individuals without restrictions assign numerals in proportion to perceived intensities of whatever is being scaled, resulting in ratio data and the possibility to describe relations among sensory magnitudes (see, e.g., Stevens, 1975; Marks, 1974). In many applied settings, however, ratio scaling procedures such as ME are impractical to use. They also lack an important aspect present in more simple category scales with verbal anchors, the possibility of obtaining direct level determinations (of, for example, what is perceived as "weak" or "strong").

The Borg CR Scales (Category-Ratio) are "level-anchored ratio scales" developed to combine the advantages of category scales with those of ratio scales. Some important principles in scale construction have been Gunnar Borg's Range model for interindividual comparisons, magnitude estimation for ratio data, choice of the most appropriate verbal anchors, and congruence between the verbal anchors and the ratio scale (G. Borg, 1982, 1998, G. Borg, & E. Borg, 2001). The latest CR scale, the Borg centiMax® Scale (0–100, also called CR100) is shown in Fig. 1. "Maximal" (100) is anchored in a previously experienced maximal perceived exertion (of, for example, lifting something so heavy so that you can just barely manage to lift it). This maximal perception constitutes a unit and the scale gives measures in centigrade of this maximum, hence the name "centiMax" (cM). The CR Scales have been validated by physiological variables and performance measures of perceived exertion, and in other sensory areas by comparison to ME, and have mainly been used to assess pain, perceived exertion, shortness of breath, muscle fatigue, etc. (see, for example, G. Borg, 1998, E. Borg, 2007, E. Borg *et al.*, 2010, Dedering, Németh, & Harms-Ringdahl, 1999; Kendrick, Baxi, & Smith, 2000; Just *et al.*, 2010). However, being general intensity scales, the Borg CR scales may be used also in the study and diagnosis of emotional variables such as those present in major depression. Because depression is diagnosed from the subjective symptoms and suffering of the patients it is important to be able to make as accurate and information-rich measurements as possible.

In a series of studies the Borg centiMax® Scale was compared to the Beck Depression Inventory (BDI, Beck *et al.*, 1961) and found to work very well. The BDI is constructed to measure major depression with 21 items rated on four-point category scales. The result showed correlations of 0.754 and 0.824 between the two scales (Borg, Magalhaes, Mörtberg, & Costa, 2016; Magalhaes, 2017).

The Patient Health Questionnaire (PHQ-9) is another popular self-report instrument to measure depressive symptoms (Kroenke, Spitzer, & Williams, 2001). The scale uses the frequency of symptoms in a four-point category scale from "Not at all" (0) to "Nearly every day" (3), summed into a total score (0–27) of depression severity (None, Mild, Moderate,

Moderately Severe and Severe) with a cut-off score of ≥ 10 for detecting cases of current Major Depressive Episode (Kroenke *et al.*, 2009). Reported sensitivity and specificity for PHQ-9≥ 10 has been high and above 80% (see, e.g., Kroenke, Spitzer, & Williams, 2001; Kroenke, Spitzer, Williams, & Löwe, 2010; Mitchell, Yadegarfar, Gill, & Stubbs, 2016). The main purpose of this study was to explore and improve the measurement of depression. For this reason, a non-clinical sample was asked to complete a well-established depression instrument that measures frequency of depressive symptoms (PHQ-9). Additionally, they were asked to respond to the same items on the Borg centiMax® Scale.

Method

Participants

A non-clinical convenience sample (n = 71, 40 men and 31 women), mean age 32 (sd = 11) years, participated. They were informed that the survey was anonymous and voluntary, for research purposes only, and that the survey could be terminated at any point. Participants that did not fulfil the entire questionnaire were excluded from the analysis.

Materials and Procedure

In an online questionnaire (administered in Survey & Report), depression was first measured with The Patient Health Questionnaire (PHQ-9) (Kroenke, Spitzer, & Williams, 2001). The PHQ-9 is a self-administered diagnostic instrument for depression consisting of nine items with responses given in relation to how often the symptoms have bothered the person the last 2 weeks. Answers are given on a 4 categories rating scale (not at all (0), several days (1), more than half the days (2), nearly every day (3)) and added into a total score from 0–27 of depression severity for None, Mild (5–9 points), Moderate (10–14 points), Moderately Severe (15–19 points), and Severe depression ($\geq$ 20 points).

The second part the full instruction for the Borg centiMax® Scale (Cr100, cM) was given, with a maximal perceived exertion as the main reference (G. Borg & E. Borg, 2001, 2016). The questions from the PHQ-9 were adapted to fit the centiMax Scale and subjects were asked how intense the feelings had been (in centigrade/cM of maximum). The sixth item on the PHQ-9 (“Feeling bad about yourself—or that you are a failure or have let yourself or your family down”) was separated into two items with the centiMax, due to it asking for two different emotions (low self-esteem and feelings of failure). The

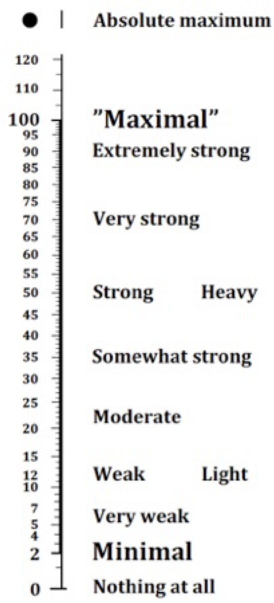


Fig. 1. The Borg centiMax Scale® (CR100, cM) used in this article ® G. Borg & E. Borg, 2001, 2016; E. Borg, 2007. See also www.borgperception.se. Printed with permission.

cM-values for these two items were then averaged so that for the analysis both scales had the same number of items.

Results

Descriptive statistics are presented in Table 1. The average PHQ-9 score for the total sample was 6.1 ($sd = 5.3$) (Mild depression) and 20 cM ($sd = 16$ cM) on the Borg centiMax Scale (just below Moderate, see Figure 1). In total, 13 persons (18%), 9 men (22.5%) and 4 women (12.9%) had PHQ-9$\geq$ 10 (moderate depression). Data for both scales was somewhat positively skewed (for PHQ-9: 1.5; for centiMax: 1.6) with positive kurtosis (for PHQ-9: 1.9; for centiMax: 3.0).

The Pearson correlation between the two scales was $r = 0.78$ ($p < 0.001$) ($r_s = 0.69, p < 0.001$). Linear regression with PHQ-9 as independent variable and centiMax gave $B_0 = 5.0$ ($p = 0.010$) and $B_1 = 2.4$ ($p < 0.001$). By using this equation, a cut-off score of 29 cM (equivalent to PHQ = 10) was calculated for Moderate depression. This corresponds to an average feeling of the intensity of the symptoms between Moderate and Somewhat strong (Fig. 1). In the total sample, 12 persons had an average centiMax $score \geq 29cM$ and 8 of these were the same persons that had PHQ-9$\geq$ 10. The average PHQ-9 and centiMax scores are shown in Table 2. The PHQ-9 had a high level of internal consistency, with a Cronbach's alpha of 0.869. For the centiMax scale, Cronbach's alpha was 0.874 (9 items) and 0.889 (10 items). To study intra- or interindividual differences a symptoms profile may be constructed. This is illustrated in Fig. 2 for two participants with the same value on the PHQ-9 (14, Moderate depression) and rather similar mean values on the centiMax (35 and 36 cM respectively). As can be seen profiles of two individuals can be rather dissimilar.

Table 1. Mean, Medians and Standard Deviations for depression measured with the PHQ-9 and the Borg centiMax® Scale.

		PHQ-9			centiMax (cM)		
Gender	*Count*	*Mean (sum)*	*Md*	*SD*	*Mean*	*Md*	*SD*
Men	40	6.6	4.0	6.1	21	16	19
Women	31	5.4	4.0	4.1	17	16	11
Total	71	6.1	4.0	5.3	20	16	16

Table 2. Number of participants with at least Moderate depression and their average scores.

	PHQ-9		centiMax (cM)	
Depression	*Count*	*Mean*	*Count*	*Mean*
None - mild	58	4.1	59	14
$\geq$ Moderate*	13	15.3	12	49

*PHQ-9$\geq$ 10 and $cM \geq 29$ respectively

Discussion

The average PHQ-9 score in this convenience sample was 6.1, and 18% had PHQ-9$\geq$ 10. This is somewhat more than the average score of 3.7 and a prevalence of 10.8% (PHQ $\geq$ 10) that was reported for the Swedish population by Johansson *et al.* (2013). Because the main purpose of this study was scale comparison, this deviance should, however, be of little consequence.

Data was somewhat positively skewed. This is not surprising because most people are not depressed, and with a non-clinical sample a somewhat restricted range of responses was to be expected. The Pearson correlation between PHQ-9 and the Borg centiMax® Scale was still high, $r = 0.78$ and the Spearman correlation obviously somewhat lower, $r_s = 0.69$. As a comparison, Kroenke *et al.* (2001) reported a test-retest correlation for PHQ-9 of $r = 0.84$ (within 48 hours). In comparison to this, 0.78 shows a very good correspondence between the two scales and the centiMax can be assumed to work at least equally well to measure depression as the PHQ-9. Obviously, a correlation of 1.0 is not what we want in this context, because of course we want the new scale to be better and contribute something more.

The internal consistency reliability was high for both scales, with a Cronbach's alpha of 0.869 for PHQ-9 and 0.874 (9 items) for centiMax. This is in agreement with what has been found previously (e.g., Kroenke *et al.*, 2001). It shows that the frequency scale of the PHQ-9 can very well be replaced by intensity estimates with the centiMax scale.

The symptom profiles (Fig. 2) illustrate some of the differences between the two scales. The two participants were chosen because they had a score above PHQ-9$\geq$ 10 and also the same PHQ-9 score (14). Sometimes the intensity of a symptom is paralleled by a high frequency of occurrence, and sometimes not. For example, the two participants rated "poor appetite or overeating" at 35 cM (solid) and 40 cM (dotted) respectively, but the first subject (solid) had this feeling (a "Somewhat strong" intensity) nearly every day. Both of them felt "Loss of self-esteem and being a failure" more than half the days, but

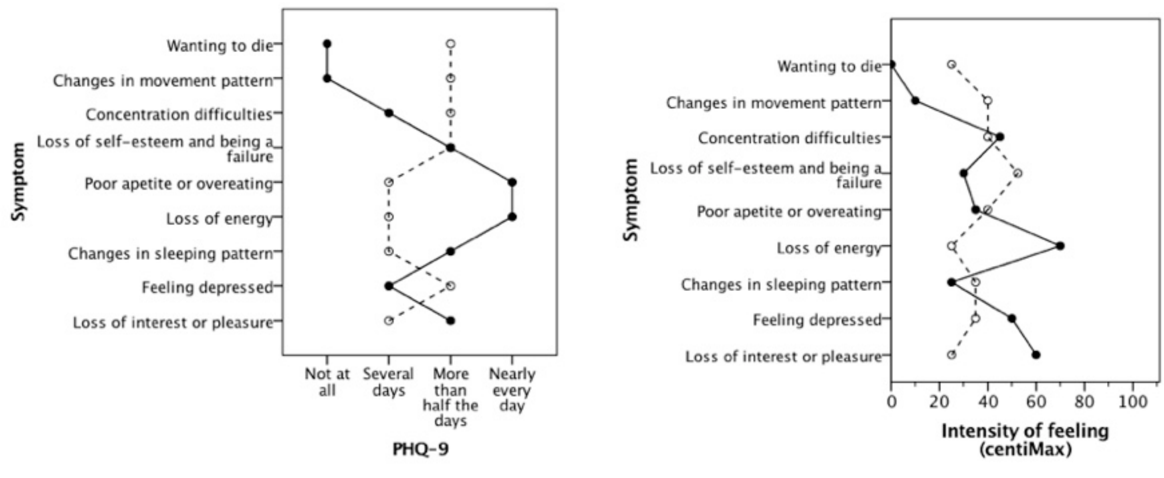


Fig. 2. Symptoms profiles for PHQ-9 (left) and centiMax (right) for two participants with the same PHQ-9 score (14, Moderate Depression) and similar mean values on the centiMax.

for one of them the feeling was almost twice as intense (1.75 times). This also illustrates another important difference: with the Borg centiMax® scale it is possible to study both the "absolute" intensity-level and how much more intense the symptoms are perceived. Table 2 shows, for example, that in the group of 12 persons who had a centiMax score above 29, the average feeling was close to "Strong" (Fig.1) and at the same time 3.5 times as strong as in the group with none to mild depression. No such information can be derived from PHQ-9.

In summery, the present study shows that using the Borg centiMax® Scale (CR100) to measure depression works very well. Of course, much research remains, but by obtaining measurements of intensity levels as well as ratio data, you can probably get interesting new insights into complex psychological phenomena such as depression.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4(561-571).

Borg, E. (2007). *On perceived exertion and its measurement.* Doctoral dissertation. Department of Psychology, Stockholm University, Sweden.

Borg, E., Borg, G., Larsson, K., Letzter, M., & Sundblad, B.-M. (2010). An index for breathlessness and leg fatigue. *Scandinavian journal of medicine & science in sports*, 20(4), 644–650.

Borg, G. (1982). A category scale with ratio properties for intermodal and interindividual comparisons. In H.-G. Geissler & P. Petzold (Eds.), *Psychophysical Judgement and the Process of Perception* (pp. 25-34). Berlin: VEB Deutscher Verlag der Wissenschaften.

Borg, G. (1998). *Borg's Perceived Exertion and Pain Scales.* Champaign: Human Kinetics.

Borg, G, & Borg, E. (2001). A new generation of scaling methods: Level anchored ratio scaling. *Psychologica*, 28, 15-45.

Borg, G., & Borg, E. (2016). The Borg CR Scales® Folder. *Methods for measuring intensity of experience.* In B. Perception (Ed.), www.borgperception.se. Hässelby, Sweden: Borg Perception.

Borg, E., Magalhaes, A., Mörtberg, E., & Costa, M. (2016, submitted). A pilot study comparing The Borg CR Scale® (centiMax®) and the Beck Depression Inventory for scaling symptoms of Depression. *Nordic Psychology.*

Dedering, A., Németh, G., & Harms-Ringdahl, K. (1999). Correlation between electromyographic spectral changes and subjective assessment of lumbar "muscle fatigue" in subjects without pain from the lower back. *Clinical Biomechanics*, 14, 103-111.

Ferrari, A., Somerville, A., Baxter, A., Norman, R., Patten, S., Vos, T., & Whiteford, H. (2013). Global variation in the prevalence and incidence of major depressive disorder: A systematic review of the epidemiological literature. *Psychological Medicine*, 43(3), 471-481. doi:10.1017/S0033291712001511

Johansson, R., Carlbring, P., Heedman, Å., Paxling, B., & Andersson, G. (2013). Depression, anxiety and their comorbidity in the Swedish general population: point prevalence and the effect on health-related quality of life. *PeerJ*, 1(6), e98. http://doi.org/10.7717/peerj.98

Just, N., Bautin, N., Danel-Brunaud, V., Debroucker, V., Matran, R., & Perez, T. (2010).

The Borg dyspnoea score: A relevant clinical marker of inspiratory muscle weakness in amyotrophic lateral sclerosis. *European Respiratory Journal*, 35(2), 353–360. http://doi.org/10.1183/09031936.00184908

Kendrick, K. R., Baxi, S. C., & Smith, R. M. (2000). Usefulness of the modified 0-10 Borg scale in assessing the degree of dyspnea in patients with COPD and asthma. *Journal Of Emergency Nursing*, 26, 216-222

Kroenke, R.L., Spitzer, J.B., & Williams (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16, 606–613.

Kroenke, K., Spitzer, R. L., Williams, J. B. W., & Löwe, B. (2010). The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. *General Hospital Psychiatry*, 32(4), 345–359.

Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B., Berry, J. T., & Mokdad, A. H. (2009). Research report: The PHQ-8 as a measure of current depression in the general population. *Journal Of Affective Disorders*, 114, 163-173.

Magalhaes, A. (2017). Três estudos usando a escala centiMax de Borg® (escala CR de Borg®, CR100, cM) para o escalonamento de sintomas depressivos. Doctoral dissertation. Department of Experimental Psychology, Institute of Psychology, University of São Paulo, Brazil.

Marks, L. E. (1974). *Sensory Processes: The New Psychophysics*. New York: Academic Press.

Mitchell, A. J., Yadegarfar, M., Gill, J., & Stubbs, B. (2016). Case finding and screening clinical utility of the Patient Health Questionnaire (PHQ-9 and PHQ-2) for depression in primary care: a diagnostic meta-analysis of 40 studies. *British Journal of Psychiatry Open*, 2, 127–138.

Montgomery S, Åsberg M. (1979). A new depression scale designed to be sensitive to change. *British Journal of Psychiatry*, 134: 382-9.

Richards, D. (2011). Prevalence and clinical course of depression: A review. *Clinical Psychology Review*, 31(7), 1117–1125. http://doi.org/10.1016/j.cpr.2011.07.004

Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. New York: Wiley.

WHO. (2016, December 7). *World Health Organization*. Retrieved March 2, 2017, from World Health Organization, http://www.who.int/topics/depression/en

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, 67(6), 361-370.

USING PSYCHOPHYSICAL SCALING, THE BORG CENTIMAX SCALE® (CR100), IN QUESTIONNAIRES TO STUDY WORK-RELATED NEEDS AND BEHAVIORS

Elisabet Borg, Johanna Andersson, and Elin Wigert
Department of Psychology, Stockholm University, SE-10691 Stockholm, Sweden
<eb@psychology.su.se>

Abstract

In test construction, thorough effort is put into formulating test items. Commonly, however, these are scaled with category or Likert scales. In this study, items from four work-related questionnaires were adapted and scaled with the Borg CR Scale® (CR100, centiMax®), a general intensity scale for level determination with ratio. Tests measured Work-related Basic Need Satisfaction, Organizational Citizenship Behavior, Perceived Organizational Support, and Procedural Justice. In their original forms, Cronbach's alpha ranged from 0.74–0.98, whereas this study obtained 0.83–0.92 (n = 30 and n = 81). Adapted to the centiMax scale, Cronbach's alpha were 0.72–0.95 ($n = 81$ and $n = 142$). For Work-related Basic Need Satisfaction ($n = 81$) and for Organizational Citizenship Behavior ($n = 30$) correlations between centiMax and Likert scales were $r = 0.79$ ($p < 0.001$) and $r = 0.69$ ($p < 0.001$). The centiMax scale worked well and can be recommended with the advantage of more interesting statistical analysis that comes with ratio data.

When constructing psychological tests, thorough effort is usually put into the formulation of test items in order to fully capture the phenomenon under study. Responses are then, however, commonly given on rather simple rating scales of an ordinal, or at the best, interval character, as, for example, a Likert scale. The purpose of this study was to show how psychophysical scaling in the form of the Borg centiMax Scale® (CR100, cM) can be applied in this context and to discuss the advantages with doing so.

A great advantage with psychophysical scales, such as magnitude estimation, is the resulting ratio data (Stevens, 1975, Luce & Krumhansl, 1988) that enables the study of ratio relations among scaled objects and phenomena. Traditionally, psychophysical scaling has been applied to the study of relations among sensations and perceptions, and psychophysical S-R-functions have been described for loudness, taste, exertion, pain, brightness, etc. (e.g., Stevens, 1975, Marks, 1974). In the psychometric test tradition of scales for personality characteristics and behaviors, the focus was usually only to study more or less of a phenomenon. If a person scores high, he or she has much of a measured characteristic, and how much might be measured in standard deviations from the average or in percentage of the population. With ratio data, one could, of course, also answer "how much more"?

The Borg CR (Category-Ratio) Scales have been developed within the psychophysical tradition with the aim to combine the advantages of verbal anchors for interindividual comparisons and level determination ("how strong?"), with those of a ratio scale for more correct descriptions of relations among perceptions ("how much stronger?"). There are several important principles underlying the scale construction. Perhaps most important, is Gunnar Borg's Range model, according to which maximal perceptual intensities (of, e.g., exertion) are set equal across individuals and "the intensity of a perception is evaluated depending upon its position in the total range from zero... to a maximal subjective intensity" (G. Borg, 1990). Other important principles are the usage of a maximal perceived exertion as a specific anchor or reference point (a "fixed star"), a reasonable size of the numerical scale that corresponds closely to the subjective dynamic range, quantitative semantics for choice of the best verbal anchors, and congruence between anchors and scale values so as to obtain ratio data similar to what is obtained with magnitude estimation (G. Borg, & E. Borg, 2001). Mainly two CR scales are now in use, the Borg CR10 Scale (Borg, G., 1982, 1998), and the Borg centiMax® Scale (also called CR100) from 0 to 100 (Fig. 1). According to the instruction, "Maximal" (100) is anchored in a previously experienced maximal perceived exertion, the feeling you have when you, for example, lift something that is so heavy so that you can just barely manage to lift it (G. Borg & E. Borg, 2016). This maximal perception constitutes a unit and the scale gives measures in centigrade of this maximum, hence the name "centiMax" (cM).

Both the CR10 and centiMax scales have primarily been developed for sensory perception, and especially for applications in physical work and effort (G. Borg, 1998; E. Borg, 2007). The scales are, however, general intensity scales and can be assumed to function equally well with all subjective variables with an intensity variation (from weak to strong). Because of the advantages of data obtained with a "level-anchored ratio scale" it would be interesting to apply a CR scale (e.g., the centiMax) to more abstract psychological constructs.

A couple of validated tests measuring work-related psychological variables studied in a bachelor thesis (Anderson & Wigert, 2017) were chosen for adaptation to the Borg centiMax Scale. The tests measured Work-related Basic Need Satisfaction, Organizational Citizenship Behavior, Perceived Organizational Support, and Procedural Justice. The tests have good reliability with Cronbach's alpha between approximately 0.74–0.98. (See, e.g., Broeck, Vansteenkiste, De Witte, Soenens & Lens, 2010; Lee & Allen, 2002; Niehoff & Moorman, 1993; Gürbüz & Mert, 2009; Eisenberger, Huntington, Hutchison, & Sowa, 1986; Eisenberger, Stinglhamber, Vandenberghe, Sucharski och Rhoades, 2002).

The purpose of Anderson and Wigert (2017) was to explain Organizational Citizenship Behavior by the other variables. The primary aim of this article was, however, to study the implementation of the Borg centiMax Scale® (CR100, cM) in the questionnaires.

Method

Three separate studies were used for data collection. The questionnaires were all administered in a web-based survey tool, SurveyMonkey®. Participants were informed about the objectives of the studies, could quit whenever they wished, and all answers were totally anonymous.

Participants

Study 1. A convenience sample of psychology students at Stockholm university, and friends of the junior authors served as participants and were reached through email and Facebook. 81 persons (64 women and 17 men) answered the whole questionnaire (after a dropout of 13 persons who did not answer all questions). Mean age was 29 years ($sd = 8$ years).

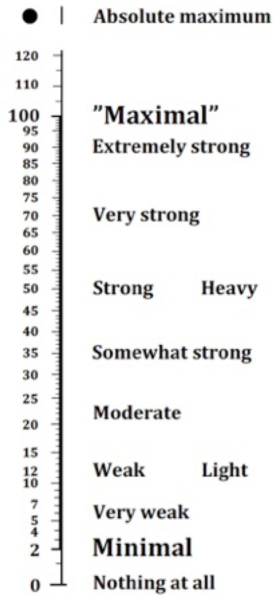


Fig. 1. The Borg centiMax Scale® (CR100, cM) used in this article ® G. Borg & E. Borg, 2001, 2016; E. Borg, 2007. See also www.borgperception.se. Printed with permission.

Study 2. The second questionnaire was after a dropout of 7 persons, answered by a convenience sample of 30 participants (17 women, 12 men and 1 unknown), from the social networks of the authors. Mean age was 35 years ($sd = 13$ years).

Study 3. The third questionnaire was also answered by a convenience sample of students and friends to the authors. In total, 189 persons started answering the questionnaire, 46 were excluded because they had not answered all questions, or worked less than 8 hours per week, one subject was found to be an outlier ($z_{residuals} < -3.3$), and 142 answers (86 women and 56 men) could be used for analysis. Mean age for these participants was 32 years ($sd = 9$ years).

Materials and Procedure

Study 1. Basic psychological needs were measured with the Work-related Basic Need Satisfaction scale (W-BNS). It consists of 18 items for the three domains of autonomy, competence and relatedness with responses given on a five-point Likert scale (1 = totally disagree to 5 = totally agree) (Broeck et al., 2010). Items were translated into Swedish by the authors. To adapt the scale for usage with the Borg centiMax Scale eight items with the highest factor loadings according to Broeck et al. (2010) and that were not reversed, were chosen. Items were also adjusted from a "disagree – agree" formulation to fit the "weak – strong" dimension of the centiMax scale. For example, an item with the former formulation was: "I am good at the things I do in my job"; and adapted to fit the centiMax: "How strong is your feeling of... being good at the things you do in your job?" Participants first answered the 18 items with the Likert scale (W-BNS(L)). Following this, they read the full instruction for the centiMax Scale, including the usage of a previously experienced maximal perceived exertion as a reference point for 100 cM. After this, they answered the remaining eight items with the centiMax Scale (W-BNS(cM)).

Study 2. The 16 items of the Organizational Citizenship Behavior Scale (OCB-S) (Lee & Allen, 2002) were first adapted to a Likert scale ranging from 1 (totally disagree) to 5 (totally agree), OCB-S(L) as well as to the centiMax scale OCB-S(cM). Participants obtained the full instruction for the centiMax scale, and then answered OCB-S(cM) followed by OCB-S(L).

Study 3. Four tests were included, adapted for the centiMax scale: the full Organizational Citizenship Behavior Scale ten items from the Work-related Basic Need Satisfaction Scale (see above); together with six items to measure Procedural Justice (Niehoff och Moorman, 1993; Gürbüz och Mert, 2009); and a scale with three items from the Survey of Perceived Organizational Support (SPOS) (Eisenberger et al., 1986; Eisenberger, et al., 2002).

Results

For each of the scales an index for each individual was calculated as the average for all questions for that person. Descriptive statistics for the scales in each of the three studies are presented in Table 1. Mean values obtained with the centiMax scale were between 49 and 65 cM. This corresponds to feelings between "strong" and "very strong" for the different psychological constructs (cf. Fig. 1). Index values for all scales were approximately normally distributed. For all scales reliability was assessed with Chronbach's alpha (see Table 1). For study 1 index values were also calculated for each of the subscales autonomy, competence and relatedness based on the 8 items that were the same for both scales (see Table 2). Fig. 2 shows individual profiles of psychological needs from 4 participants in Study 1. The participants were chosen for the sake of illustration. The Pearson correlation between W-BNS(L8) and W-BNS(cM8) was $r = 0.79$ ($p < 0.001$) (Study 1), and between OCB-S(L) and OCB-S(cM), $r = 0.69$ ($p < 0,001$) (Study 2).

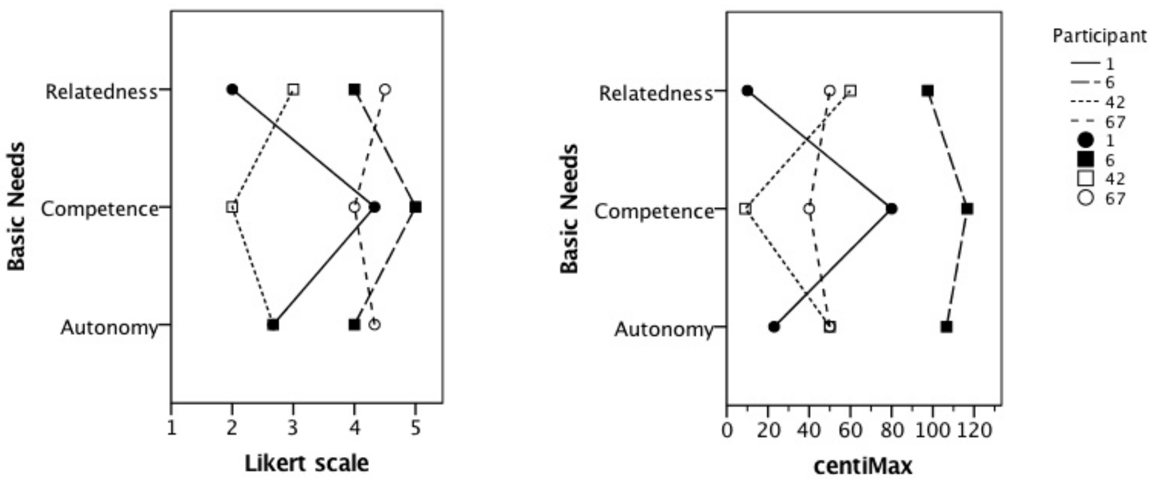


Fig. 2. Individual profiles of psychological needs from 4 participants in Study 1 with the Likert scale (left) and centiMax (right).

Table 1. Mean values, standard deviations (SD), Skewness, Kurtosis, and Cronbach's alpha (Alpha) for the different psychological constructs and scales used in each study.

Study (N)	Construct	Scale	Mean	SD	Skewness	Kurtosis	Alpha
1 (81)	Work-related basic need satisfaction	W-BNS(L) 18 items	3.8	0.60	-0.66	-0.19	0.86
		W-BNS(L) 8 items	3.8	0.70	-0.60	-0.43	0.79
		W-BNS(cM) 8 items	65	22	-0.26	-0.83	0.92
2 (30)	Organizational citizenship behavior	OCB-S(L) 16 items	3.7	0.53	-0.63	1.15	0.84
		OCB-S(cM) 16 items	58	17	0.27	-0.68	0.94
3 (142)	Work-related basic need satisfaction	W-BNS(cM) 10 items	59	18	-0.07	-0.35	0.89
	Organizational citizenship behavior	OCB-S(cM) 16 items	58	20	-0.03	-0.23	0.95
	Procedural justice	PJ (cM) 6 items	51	22	0.12	-0.06	0.92
	Perceived organizational support	SPOS (cM) 3 items	49	22	0.18	-0.41	0.88

Table 2. Mean values for the subscales of W-BNS based on 8 items (Study 1, $N = 81$)

W-BNS (no. of items):	Autonomy (3)	Competence (3)	Relatedness (2)
Likert scale (1–5)	3.5	4.2	3.7
centiMax	60	71	63

Discussion

The aim of this article was to implement the Borg centiMax® Scale (CR100, cM) and study its value in questionnaires on work-related needs and behaviors. The centiMax scale is a "level-anchored ratio scale", meaning that obtained data gives information about both "absolute" intensity levels and ratio relations (G. Borg, & E. Borg, 2001). For two tests (W-BNS and OCB) the items were scaled both with a Likert scale and with the centiMax and the correlations obtained were 0.79 and 0.69 respectively. This show that the reformulation of items to be measured in centiMax according to a variation in perceived intensity is valid. Table 1 also shows that reliability measured as internal consistency with Cronbach's alpha was high for the Likert scale and in agreement with values previously obtained (e.g. Broeck et al., 2010). With the centiMax scale even somewhat higher values were obtained. This shows that rephrasing items and using them together with centiMax gives at least equally reliable measurements.

The unique value of centiMax is, however, best illustrated by Table 2 and Fig. 2. While the profiles are similar, data obtained with the Likert scale only give information about “much” or “little”, “more” or “less” satisfaction of basic psychological needs. In reality, what the mean values in Table 2 tell us is only that the group on average most agreed to the items on competence, a little less to items on relatedness and even a little less to items on autonomy, but that they on average agreed more than disagreed. The cM-values show that the intensity in their feelings was on average between 60 and 71 cM, or percent of “Maximum”, and close to a “very strong” feeling (Fig. 1). Fig. 2 illustrates this further. For example: Participant No. 1 does not only feel “a lot more” competent than related or autonomic, but feels between very and extremely competent and 8 times as competent as related (below weak) and about 5 times as competent as autonomic (moderately). Participant No. 6 shows the value of allowing responses above 100 cM (which is necessary to avoid a ceiling effect and may also result from using “maximal exertion” as a reference). Participant No. 42 illustrates the value when screening for a specific purpose, for example to find employees in need of competence development. And, finally, participant 67 points to the fact that someone might “agree” to a rather large extent and still have only “somewhat strong” feelings.

In summary, the present study shows that the use of a level-anchored ratio scale, i.e., the Borg centiMax® Scale (CR100) instead of a Likert scale works very well. Of course, much research remains, but measurements of intensity levels as well as ratio data may well contribute to new insights into complex psychological phenomena. For example, Chen et al., (2015) studied basic psychological need satisfaction across four cultures and one conclusion was that “satisfaction of each of the three needs was found to contribute uniquely to the prediction of well-being” regardless of cultural background. With data in centiMax one could perhaps also answer questions on how big this contribution might be.

References

Anderson, J., & Wigert, E. (2017). Grundläggande psykologiska behov och anställdas benägenhet att “go the extra mile”. (Unpublished bachelor thesis. Swedish.) Department of Psychology, Stockholm University, Sweden.

Borg, E. (2007). *On Perceived Exertion and its Measurement*. Doctoral dissertation. Department of Psychology, Stockholm University, Sweden.

Borg, E., Borg, G., Larsson, K., Letzter, M., & Sundblad, B.-M. (2010). An index for breathlessness and leg fatigue. *Scandinavian Journal of Medicine & Science in Sports*, 20(4), 644–650.

Borg, G. (1982). A category scale with ratio properties for intermodal and interindividual comparisons. In H.-G. Geissler & P. Petzold (Eds.), *Psychophysical Judgement and the Process of Perception* (pp. 25–34). Berlin: VEB Deutscher Verlag der Wissenschaften.

Borg, G. (1990). A General Model for Interindividual Comparison. In W. J. Baker, M. E. Hyland, R. van Hezewijk, & S. Terwee (Eds.), Recent Trends in Theoretical Psychology: Proceedings of the Third Biennial Conference of the *International Society for Theoretical Psychology*, April, 17–21, 1989 (pp. 439–444). New York, NY: Springer US. http://doi.org/10.1007/978-1-4613-9688-8_44

Borg, G. (1998). *Borg's Perceived Exertion and Pain Scales*. Champaign: Human Kinetics.

Borg, G, & Borg, E. (2001). A new generation of scaling methods: Level anchored ratio scaling. *Psychologica*, 28, 15–45.

Borg, G., & Borg, E. (2016). The Borg CR Scales® Folder. *Methods for Measuring Intensity of Experience*. In B. Perception (Ed.), www.borgperception.se. Hässelby, Sweden: Borg Perception.

Broeck, A., Vansteenkiste, M., De Witte, H., Soenens, B. & Lens, W. (2010). Capturing autonomy, competence and relatedness at work: Construction and initial validation of the Work-related Basic Need Satisfaction scale. *Journal of Occupational and Organizational Psychology*, 83, 981–1002.

Chen, B., Vansteenkiste, M., Beyers, W., Boone, L., Deci, E. L., Van der Kaap-Deeder, J., ⋯ Verstuyf, J. (2015). Basic psychological need satisfaction, need frustration, and need strength across four cultures. *Motivation and Emotion*, 39(2), 216–236. http://doi.org/10.1007/s11031-014-9450-1

Eisenberger, R., Huntington, R., Hutchison, S. & Sowa, D. (1986). Perceived organizational support. *Journal of Applied Psychology*, 71, 500–507.

Eisenberger, E., Stinglhamber, F., Vandenberghe, C., Sucharski, I. & Rhoades, L. (2002). Perceived supervisor support: contributions to perceived organizational support and employee retention. *Journal of Applied Psychology*, 87, 565–73.

Gürbüz, S. & Mert, I.S. (2009). Validity and reliability testing of organizational justice scale: An empirical study in a public organization. *Review of Public Administration*, 42, 117–139.

Jones, D. M., Miles, C., & Page, J. (1990). Disruption of proofreading by irrelevant speech: Effects of attention, arousal or memory? *Applied Cognitive Psychology*, 4, 89–108.

Lee, K. & Allen, N.J. (2002). Organizational Citizenship Behavior and Workplace Deviance: The Role of Affect and Cognitions. Journal of Applied Psychology, 87, 131–142.

Licklider, J. C. R., & Miller, G. A. (1951). The perception of speech. In S. S. Stevens (Ed.), *Handbook of Experimental Psychology* (pp. 1040–1074). New York: John Wiley.

Luce, R. D., & Krumhansl, C. L. (1988). Measurement, scaling, and psychophysics. *Stevens' Handbook of Experimental Psychology*, 1, 3–74.

Marks, L. E. (1974).*Sensory Processes: The New Psychophysics*. New York: Academic Press.

Niehoff, B.P. & Moorman, R.H. (1993). Justice as a Mediator of the Relationship Between Methods of Monitoring and Organizational Citizenship Behavior. *Academy of Management Journal*, 36, 527–556.

Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. New York: Wiley.

A CNV ANALYSIS OF THE DISCRIMINATION OF VERY SHORT TIME INTERVALS

Takako Mitsudo[1], Naruhito Hironaga[1], and Simon Grondin[2]
[1]*Department of Clinical Neurophysiology, Neurological Institute Faculty of Medicine Graduate School of Medical Sciences, Kyushu University*
[2]*École de Psychologie, Université Laval*

We compared the contingent negative variation (CNV) in short standard (200 ms) and long standard (700 ms) duration conditions to replicate and extend Pfeuty et al. (2003)'s finding about the discrimination of time intervals. Sixteen participants memorized standard duration at the beginning of each trial (Direct comparison task) or block (Matching-to-sample task) in order to compare it to each of 5 comparison stimuli. EEG was recorded continuously from 32 scalp sites. Behavioral results showed that there are smaller individual differences in the Direct comparison task than in the Matching-to-sample task. EEG signals revealed that in the 700-ms condition for both tasks, CNV peak amplitudes increased as a function of the comparison stimulus duration. In the 200-ms condition for Direct comparison task, however, CNV developed only with longer comparison stimulus durations (Fig. 1). The results suggest that CNV does not reflect perceived duration of time in shorter time intervals.

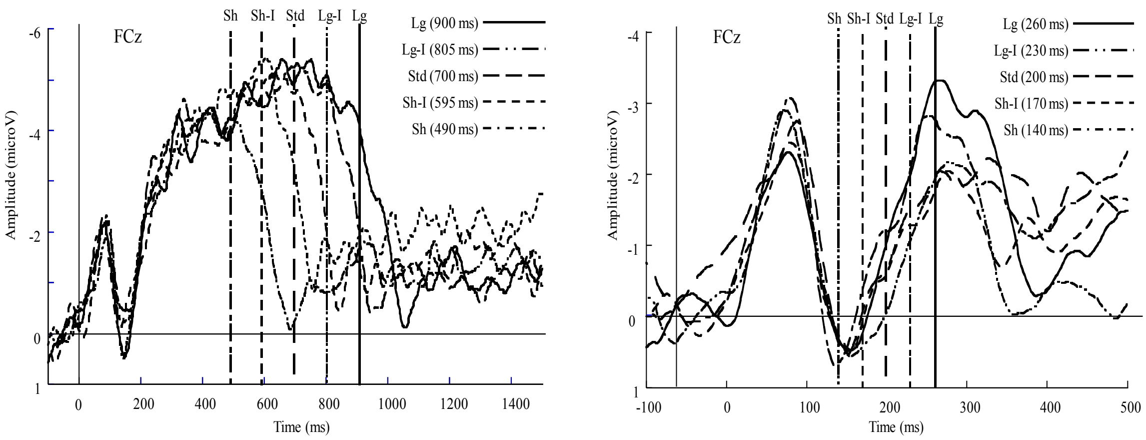


Fig. 1. ERP results during Direct comparison task. Left: 700-ms condition. Right: 200-ms condition. In 700-ms condition, CNV increased as a function of the stimulus duration. In 200-ms condition, however, CNV was observed only with longer durations (230 and 260 ms).

References

Pfeuty M, Ragot R, Pouthas V. (2003). When time is up: CNV time course differentiates the roles of the hemispheres in the discrimination of short tone durations. *Exp Brain Res,* 151(3): 372-9.

IS THE ACOUSTIC LANGUAGE UNIVERSAL REALLY UNIVERSAL?

Yoshitaka Fujii, Yoshitaka Nakajima, Takuya Kishida, and Takeharu Seno
Faculty of Design, Kyushu University, Fukuoka, Japan

The auditory system is assumed to have a property common to the speakers of all languages. Therefore, different languages must have common acoustic properties. Ueda and Nakajima [(2017). *Sci. Rep.*, 7, 42468] showed a common property across different languages: Common power fluctuation factors in sets of critical-band filters appeared following principal component analysis and varimax rotation. Although they were inclined to think that they discovered an acoustic language universal, the languages analyzed in their study were limited to West-European and East-Asian languages. We thought it would be productive to examine whether their idea could be extended to other languages covering different areas of the world. We performed almost the same analysis for another speech database, in which 21 languages were included. Some languages were common to the languages as in Ueda and Nakajima's study, and thus the replicability of their results was also checked in the present study. The obtained power fluctuation factors of all the 21 languages showed very similar properties as in Ueda and Nakajima's study. It is to be noted that the results were similar among speakers, who manifested fundamental frequencies in considerably different ranges—this was revealed by smoothing the spectra to be analyzed. (Supported by the JSPS.)

EFFECTS OF SOUND MARKER DURATIONS ON THE PERCEPTION OF INTER-ONSET TIME INTERVALS: A STUDY WITH REAL AND SYNTHESIZED PIANO SOUNDS

Emi Hasuo
School of Information Environment, Tokyo Denki University
2-1200 Muzai Gakuendai, Inzai-shi 270-1382, Chiba, Japan

Yoshitaka Nakajima
Department of Human Science/Research Center for Applied Perceptual Science
Kyushu University, 4-9-1 Shiobaru, Minami-ku, Fukuoka 815-8540, Japan

Michiko Wakasugi and Takuya Fujioka
Human Science Course, Kyushu University
4-9-1 Shiobaru, Minami-ku, Fukuoka 815-8540, Japan

<emi8513@yahoo.co.jp, nakajima@design.kyushu-u.ac.jp>

Perceiving the duration between the onsets of two successive sounds is the basis of rhythm perception in music. Our previous studies showed that a time interval marked by the onsets of two successive sounds tends to be perceived as longer when the second sound is lengthened. However, the sound markers in these studies, i.e. pure tone bursts with steady amplitude, may have been too simple to be related to real music. The present study examined whether this phenomenon would appear when the time interval was marked by piano sounds with complex spectral and temporal structures. In Experiment 1, three types of sounds were employed: recorded piano sounds (P), synthesized sounds with seven frequency components that simulated the harmonic and temporal aspects of the recorded piano sounds (S7), and pure tones with the temporal envelope of the piano sounds (S1). For each type of sounds, there were three durations, 40, 100, and 180 ms. These sounds marked time intervals of 120, 240, or 360 ms. The subjective durations of these time intervals were measured using the method of adjustment. Data obtained from 10 participants showed that lengthening the second marker increased the subjective duration of the interval, as in the previous studies, but only when the time interval was 120 ms and the first marker was 100 ms. In Experiment 2, only S7 sounds were used. No effect of sound durations appeared. In Experiment 3, only P sounds were used. Lengthening the second marker increased the subjective duration of the time interval, but only when the time interval was 360 ms. Summarizing the results, time intervals were perceived to be longer when the second sound was lengthened, but only in limited cases. The effect of sound durations observed in the present experiments, with complex sounds that could be observed in music, seemed weaker than in our previous studies, in which the effect of the second sound duration appeared stably for all time intervals. The weakening of the sound duration effect may be due to the decaying temporal envelopes of the present sounds.

ERP CORRELATES OF SENSORY FACILITATION AFTER EXPOSURE TO DELAYED AUDITORY FEEDBACK

Yoshimori Sugano[1], Jeroen Stekelenburg[2], Frouke de Schipper[2], and Jean Vroomen[2]
[1] *Department of Industrial Management, Kyushu Sangyo University, Japan*
[2] *Department of Cognitive Neuropsychology, Tilburg University, the Netherlands*

Perception of synchrony between a voluntary action and the sensory feedback thereof is changed after a short exposure phase to delayed auditory feedback (DAF). Auditory reaction times (RTs) are also shorter after exposure to DAF, suggesting that auditory processing is enhanced. Here, we examined the neural correlates of this speed-up. Participants were exposed to either delayed (150 ms) or subjectively-synchronous (50 ms) auditory feedback while they pressed a mouse at a constant pace. Immediately thereafter, they performed a simple auditory RT task while electroencephalogram (EEG) was recorded. The RT-distribution, fitted by an ex-Gaussian function, showed that the tau parameter (covering the tail) was significantly smaller after exposure to DAF than synced feedback (Fig.1A). The EEG-data showed that P1 and LPP component of ERP were significantly changed by DAF (Fig.1B, C, D). These results indicate that exposure to DAF causes auditory facilitation that can be attributed to both early sensory and late attentional processing stages.

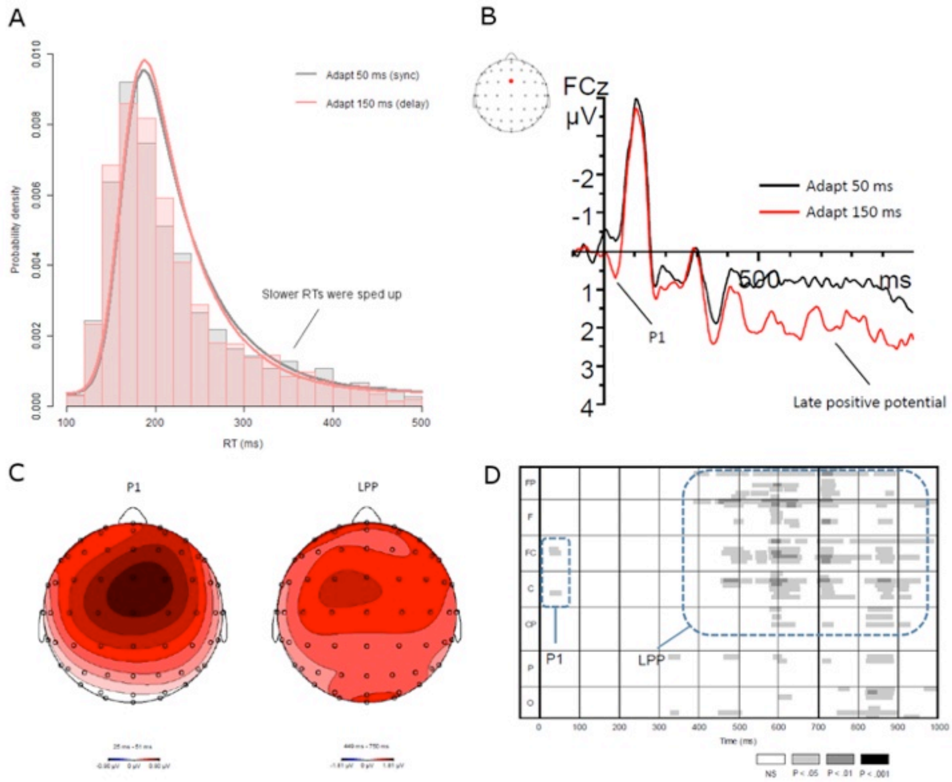


Fig. 1. (A) Group-averaged RT distribution fitted with an ex-Gaussian function, (B) Tone-locked ERPs, (C) Scalp potential maps of the difference wave for P1 (left) and LPP (right), (D) Point-wise *t*-tests on the difference wave between sync (50 ms) and DAF (150 ms).

SPECIFICITY OF PERCEPTUAL LEARNING IN AN AUDITORY INTERVAL CATEGORIZATION TASK DEPENDING ON THE DECISION RULE

Florian Kattner
Institute of Psychology, Technische Universität Darmstadt
Alexanderstraße 10, 64289 Darmstadt, Germany
<kattner@psychologie.tu-darmstadt.de>

Abstract

Performance in most perceptual tasks is known to improve considerably as a result of practice, but these learning effects are often specific to the stimuli presented during training. Only recently generalization was found to depend on whether the training task promoted the acquisition of stimulus-specific or more general perceptual decision rules. In this study, participants were trained on an auditory interval categorization task with durations drawn from uniform distributions (350–650 ms or 1100–1400 ms). Half of the participants were trained with a specific category boundary (500 or 1250 ms) for 3200 trials, whereas others were trained with a boundary that shifted on every trial. Both tasks yielded significant perceptual learning, as indicated by reduced discrimination thresholds for the trained durations. However, these improvements were specific to the trained durations when the training task could be solved by learning stimulus-specific decision rules, whereas partial transfer was observed with shifting category boundaries.

Discrimination performance of human observers is known to improve as a function of practice in most visual and auditory psychophysical tasks. However, this improvement in perceptual tasks is often highly specific to the exact stimuli and tasks that were presented during training (Ahissar & Hochstein, 1997). Only recently, research has identified some of the conditions that yield more transfer of perceptual learning to new stimuli. It has been shown, for instance, that transfer in a visual orientation discrimination task depends on whether the rules that are acquired during training can be applied to the transfer task (Green, Kattner, Siegel, Kersten, & Schrater, 2015). In this study, perceptual learning was highly specific when participants learned to discriminate orientations (between 30° and 60°) relative to a constant reference (i.e., 45°, suggesting a rule that is not applicable to new orientations), whereas transfer to new orientations (120°–150°) was observed when the reference varied from trial to trial during training. Compared to the visual domain, much less research has been dedicated to auditory perceptual learning and transfer. Perceptual learning in an auditory temporal interval discrimination task, for instance, was shown to be specific to the trained interval durations, whereas some transfer occurred to stimuli of different frequencies (e.g., Wright, Buonomano, Mahncke, & Merzenich, 1997). Interestingly, more transfer was found when participants were trained on two randomly selected standards, as compared to training with a single standard (Karmarkar & Buonomano, 2003). This seems to suggest that transfer in an auditory interval discrimination task may also depend on the generality of the perceptual decision rules that are acquired during training. To test this prediction, participants of the present study were trained on an auditory interval discrimination task with either a constant reference interval (featuring highly specific decision rules) or a reference interval that varied on every training trial (requiring the acquisition of more general decision rules). Transfer was assessed with regard to stimuli from a different range of temporal intervals.

Method

Participants

A total of 26 student participants (22 women; $M_{age} = 23.5$ years; $SD_{age} = 6.7$ years) were recruited at the campus of Technische Universität Darmstadt. Participants were compensated with course credit. The data of five participants were not included in the analysis due to extremely poor performance in the training task (average interval discrimination thresholds > 200 ms), thus leaving $N = 21$ participants for the analysis. Participants were randomly assigned to a training schedule with constant ($n = 12$) or varying reference intervals ($n = 9$).

Apparatus and Stimuli

The experiment was conducted in a single-walled sound attenuated listening booth (International Acoustics Company). Stimulus presentation and response registration were programmed in Matlab (Mathworks, Natick, MA, USA) using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). Sounds were generated in Matlab, D/A converted by an external sound card (RME multiface II) with 16 bits at 44.1 kHz, passed through a headphone amplified (Behringer HA 800 Powerplay PRO-8), and presented via headphones (Beyerdynamics DT-990, 250 Ohm). Visual information was displayed on 17" LCD monitor.

Procedure

The entire experiment took five days and consisted of a pre-test, a training and a post-test phase. Day 1 contained an initial 100-trial pre-test block with the transfer intervals and a 600-trial block of training (the first 100 trials of the training block will be considered the pre-test block for the trained intervals). Days 2 through 4 each consisted of 700 trials of training. Day 5 consisted of an additional 500-trial block of training, followed by a 100-trial post-test block with transfer intervals and another 100-trial post-test block with the trained intervals.

Each trial started with a 500-ms presentation of a fixation cross in the center of the screen. Then a burst of white noise was presented via headphones for a particular duration (including 10 ms cosine shaped rise- and fall-times) and together with a reference duration displayed within a text message on the screen (e.g., "Shorter or longer than 500 ms?"). Participants were then asked to press the "F" key when the duration of the burst of noise was shorter than the reference, and the "J" key when it was longer. Visual feedback ('Correct!' or 'Wrong!' in green or red font, respectively) was presented on the screen immediately after the response for 750 ms.

On each trial, the duration of the sound was drawn randomly from a uniform distribution. Some participants ($n = 11$) were trained on short durations (350–650 ms) and transfer was assessed with regard to long transfer durations (1100–1400 ms), whereas the remaining participants were trained on the long durations and tested on the short durations (n = 10). Moreover, for some participants, the reference interval was constant throughout the training (500 or 1250 ms, respectively), whereas it shifted on every trial

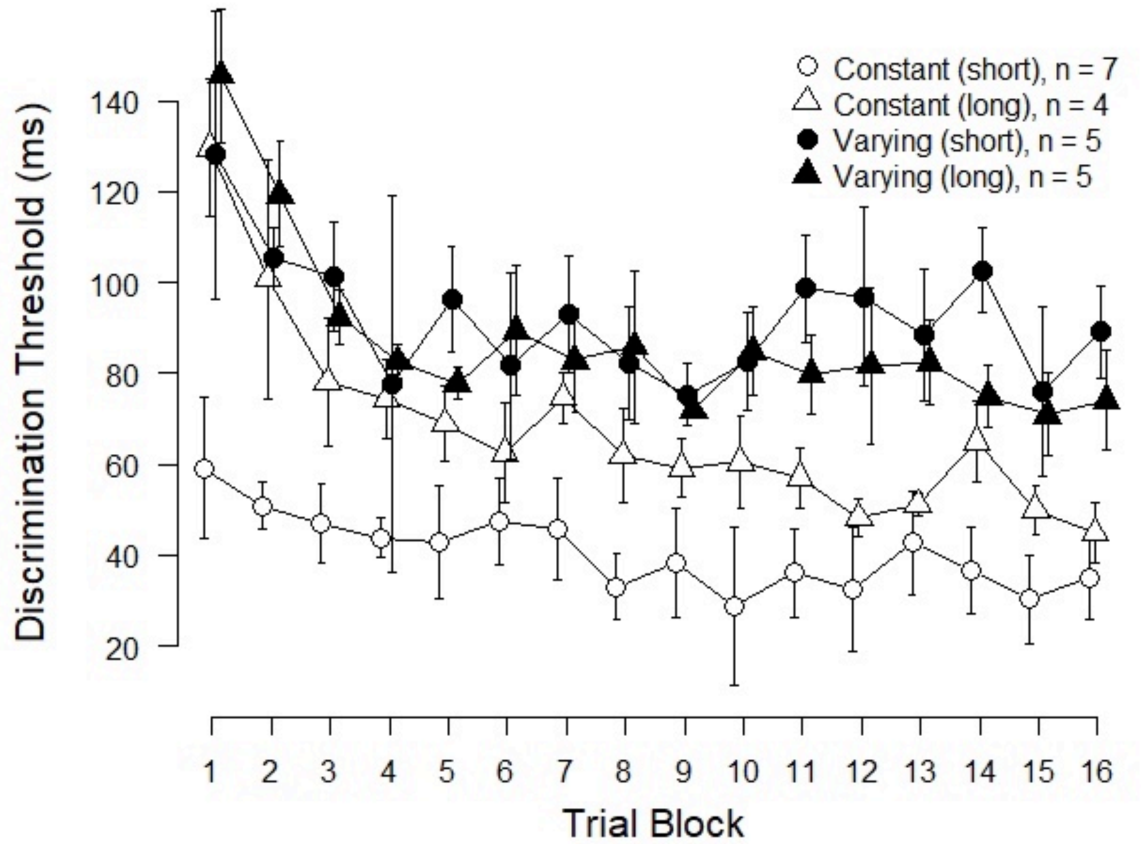


Fig. 1. Median interval discrimination 79% thresholds as a function of the training block (200 trials each) measured for participants trained with short (350–650 ms) or long (1100–1400 ms) auditory intervals using either a constant or varying reference intervals. Error bars represent interquartile ranges.

for the remaining participants (randomly drawn from 350–650 ms or 1100–1400 ms, respectively). The reference was constant for all groups during the pre- and post-test blocks with the transfer intervals as well as during the post-test block with the trained intervals.

Results and Discussion

Perceptual Learning

Temporal interval discrimination thresholds were calculated by fitting logistic psychometric functions to the participants' responses as a function of the difference between the presented interval duration and the reference. The estimated 79% thresholds decreased throughout the 3200 trials of training with short and long reference intervals and with both constant and variable reference intervals (see Fig. 1). A 16 (training block) $\times$ 2 (reference: constant vs. variable) $\times$ 2 (training interval: short vs. long) mixed-factors ANOVA with training block as a repeated measures factor revealed a significant main effect block, $F(14, 238) = 8.58$; $p < .001$; ${\eta^2}_G = 0.19$, indicating perceptual learning across all four groups. There was also a significant block $\times$ training interval

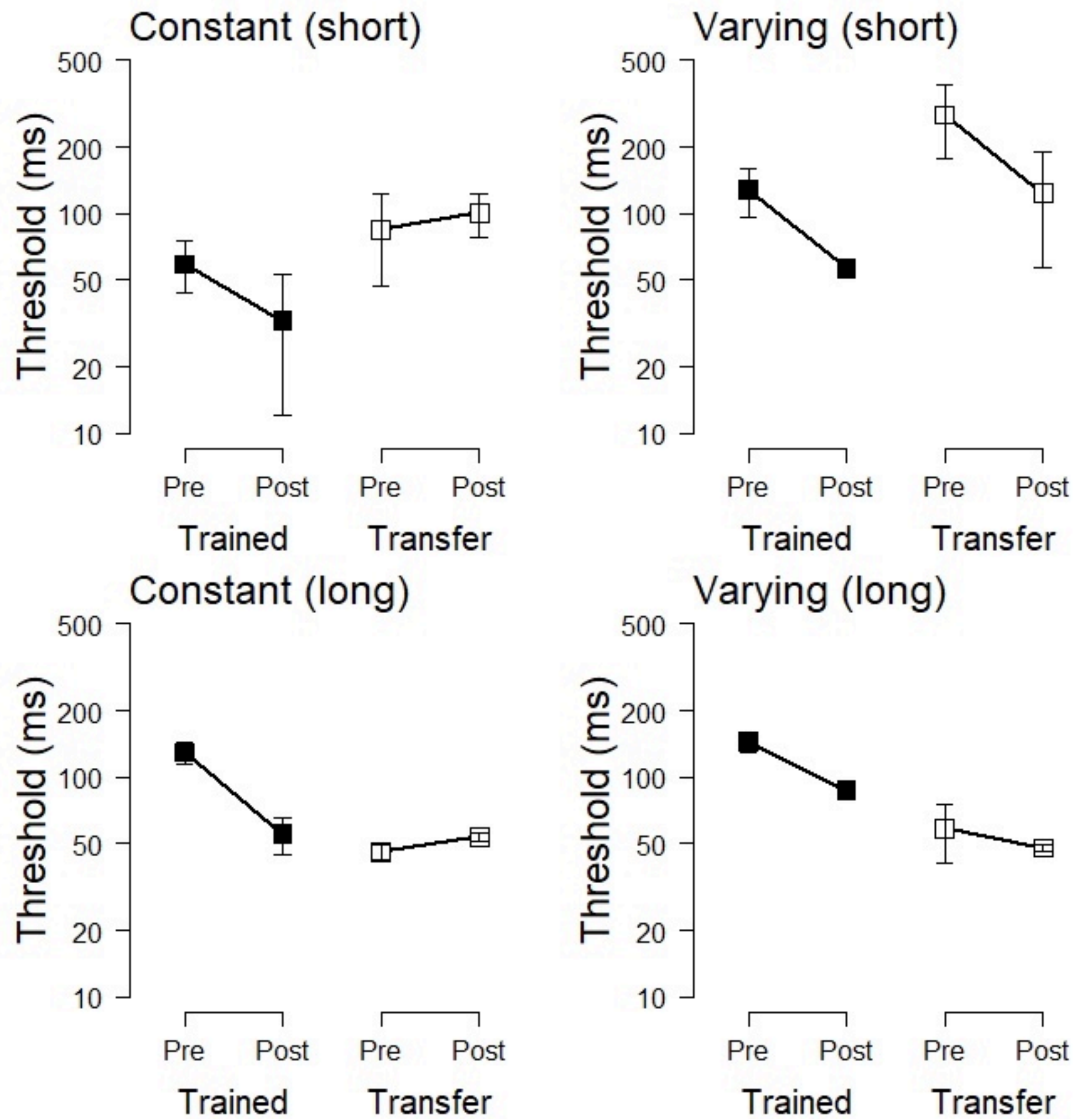


Fig. 2. Median interval discrimination thresholds (79%) obtained for the four training groups (short or long training intervals to be discriminated from a constant or varying reference) in the pre- and post-tests with the trained and non-trained (transfer) intervals. Error bars represent interquartile ranges.

interaction, $F(14, 238) = 3.39$; $p < .001$; $\eta^2{}_G = 0.08$, suggesting that greater improvements were obtained with the long ($\Delta threshold = 57.2$ ms) than with short intervals ($\Delta threshold = 17.4$ ms). Moreover, the analysis revealed a main effect of reference, $F(1, 17) = 25.49; p < .001$; $\eta^2{}_G = 0.45$, with lower overall thresholds obtained with constant than with varying references intervals, as well as a training interval × reference interaction, $F(1, 17) = 5.39$; $p = .033$; $\eta^2{}_G = 0.15$. No other effects reached significance ($F < 1$)

For the trained interval durations, discrimination thresholds also differed between the pre- and post-test blocks (see Fig. 2). A 2 (test: pre vs. post) × 2 (reference: constant vs. variable) × 2 (training interval: short vs. long) mixed-factors ANOVA with test as a repeated-measures factor revealed a significant main effect of test, $F(1, 17) = 18.74$; $p < .001$; $\eta^2{}_G = 0.34$, as well as significant main effects or reference, $F(1, 17) = 13.94$; $p = .002$; $\eta^2{}_G = 0.30$, and training interval $F(1, 17) = 13.09$; $p < .002$; $\eta^2{}_G = 0.29$. There was also a significant 3-way interaction, $F(1, 17) = 5.47$; $p = .03$; $\eta^2{}_G = 0.13$, but no other effects ($F < 1.31$).

Transfer of Learning

To assess transfer of learning, interval discrimination performance was contrasted between the pre- and post-tests with the interval durations that were not presented during training. Discrimination thresholds with these transfer durations are illustrated in Fig. 2. In contrast to the trained intervals, a 2 (test: pre vs. post) × 2 (reference: constant vs. variable) × 2 (training interval: short vs. long) ANOVA on transfer interval discrimination thresholds revealed no main effect of test, $F(1, 17) = 0.06$; $p = .80$. It can be seen that some threshold decrements were observed in participants who were trained with varying references, whereas performance did clearly not improve in participants who were trained with a constant reference. However, the analysis revealed no significant test × reference, $F(1, 17) = 0.02$; $p = .88$, indicating that performance changes with regard to the untrained durations did not differ between the two types of training schedules. There were again significant main effects of reference, $F(1, 17) = 7.87$; $p = .012$; $\eta^2{}_G = 0.14$, and training duration, $F(1, 17) = 15.97$; $p < .001$; $\eta^2{}_G = 0.24$, as well as a reference × training duration interaction, $F(1, 17) = 4.81$; $p < .042$; $\eta^2{}_G = 0.08$, but no other effects.

More observers in the varying reference groups showed decrements in discrimination thresholds from pre- to post-test for the transfer durations (6 decrements vs. 4 increments), whereas the majority of participants in the constant reference groups showed threshold increments (4 decrements and 7 increments). However, a χ^2 test did not reveal significant group differences in relative frequencies of threshold decrements, $\chi^2(3) = 3.10$; $p = .38$. Taken together, there is no evidence for a transfer of auditory interval discrimination learning to new interval durations when a single reference interval was presented during training, permitting participants to learn highly specific perceptual decision rules. However, even with varying reference intervals during training, requiring the acquisition of more general perceptual decision rules, there is only a small trend but no significant evidence for transfer to untrained intervals. This seems to suggest that there is a discrepancy between visual and auditory perceptual learning in terms of the decision rules that yield transfer (cf.; Green et al., 2015).

Acknowledgements

I am grateful to my undergraduate research assistants Michelle Benkel, Stefan Coels, Corinne Orlemann, and Alexandra von Preuschen who recruited the participants and conducted the psychophysical experiments.

References

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*. `http://doi.org/10.1038/387401a0`

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 443–446. `http://doi.org/10.1163/156856897X00357`

Green, C. S., Kattner, F., Siegel, M. H., Kersten, D., & Schrater, P. R. (2015). Differences in perceptual learning transfer as a function of training task. *Journal of Vision*, 15(10), 5. http://doi.org/10.1167/15.10.5

Karmarkar, U. R., & Buonomano, D. V. (2003). Temporal specificity of perceptual learning in an auditory discrimination task. *Learning & Memory*, 10(2), 141–147. http://doi.org/10.1101/lm.55503

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442. http://doi.org/10.1163/156856897X00366

Wright, B. A., Buonomano, D. V, Mahncke, H. W., & Merzenich, M. M. (1997). Learning and generalization of auditory temporal-interval discrimination in humans. *Journal of Neuroscience*, 17(10), 3956–3963. http://www.jneurosci.org/content/17/10/3946

THE PERCEPTION OF AUDITORY ICONS BY JAPANESE DRIVERS

João Paulo Cabral[1], Yoshitaka Nakajima[2], Kazuo Ueda[2], and Gerard B. Remijn[2]
[1]*Human Science International Course, Graduate School of Design*
Kyushu University, Fukuoka, Japan
[2]*Department of Human Science/Research Center for Applied Perceptual Science*
Kyushu University, Fukuoka, Japan
`<3ds17001e@s.kyushu-u.ac.jp, nakajima@design.kyushu-u.ac.jp`
`ueda@design.kyushu-u.ac.jp, remijn@design.kyushu-u.ac.jp>`

Many studies in the automotive field have focused on the perception and recognition of auditory warnings in various driving situations. Different from the traditional auditory warnings (e.g., vocal and abstract alarms), auditory icons are characterized by giving brief, nonverbal information through snippets of sounds from actions or events they refer to. Conveying accurate information is one of the most important attributes expected of an auditory icon. The chosen sound(s) should represent the target referent perfectly. Previous studies have shown, however, that different variations of auditory icons have been used to represent the same referent, and it is not yet understood whether drivers from different cultural backgrounds interpret various auditory icons similarly.

Here we present a preliminary study about the perception and recognition of Japanese listeners of auditory icons originally used in research in Europe and the U.S.A. During the investigation, two groups were formed. Participants in both groups listened to the same set of 15 auditory icons and were asked to describe the sounds they heard using no more than two nouns and two adjectives. Before listening, one group was informed that the auditory icons concerned vehicle driving and traffic in general, while the other group did not receive any directional instructions.

The results showed that there were no large group differences in the descriptions, but, as expected, the group in which listeners were informed that the auditory icons concerned vehicle driving and traffic was better able to describe traffic-related icons, such as a “tire skid”. Overall, the auditory icons with ubiquitous sounds (sirens, bells, voices) were described similarly by most listeners. Auditory icons with less well-known sounds were often associated with events or objects in the listener’s environment, reflecting the listener’s cultural background and experience, or described according to the physical properties of the sounds. Although used in previous research in Europe and the U.S.A, when asked to describe what they heard the Japanese listeners employed in this study did not describe all auditory icons in similar terms or in terms they were intended to represent. These preliminary results demonstrate that when sounds in auditory icons are not presented with the appropriate physical characteristics, describing the icons and the eventual universal interpretation of the correct referents of the icons will be difficult. The results open questions about how to make auditory icons more intelligible.

PERCEPTUAL RESTORATION OF INTERRUPTED LOCALLY TIME-REVERSED SPEECH

Kazuo Ueda[1], Nozomi Inui[2], Kaisei Shiraki[2], Valter Ciocca[3], Yoshitaka Nakajima[1], and Gerard B. Remijn[1]

[1]*Department of Human Science/Research Center for Applied Perceptual Science Kyushu University, Fukuoka 815-8540, Japan*

[2]*Department of Acoustic Design, Faculty of Design, Kyushu University Fukuoka 815-8540, Japan*

[3]*School of Audiology and Speech Sciences, the University of British Columbia Vancouver, BC, V6K 1X3 Canada*

`<ueda@design.kyushu-u.ac.jp>`

Locally time-reversed speech [Ueda et al. (2017). *Sci. Rep.*, 7: 1782], i.e., segmented speech with time-reversal in each segment, was interrupted by alternately replacing speech segments with silent gaps or with noise to study effects of perceptual restoration. The intelligibility of original locally time-reversed speech with 60-ms segments was about 80%. When speech segments were alternately replaced by silent gaps, however, the intelligibility dropped sharply to 10%. Our preliminary observation showed that a small but significant gain in intelligibility, about 20%, was observed by filling the silent gaps with noise (n =10; Fig. 1). A follow-up study of this observation is now undertaken, to elucidate whether the gain in intelligibility can be related to the multi-time resolution model proposed by Poeppel et al. [e.g., Giraud and Poeppel (2012). *Nature Neurosc.*, 15, 511-517]. [Work supported by JSPS.]

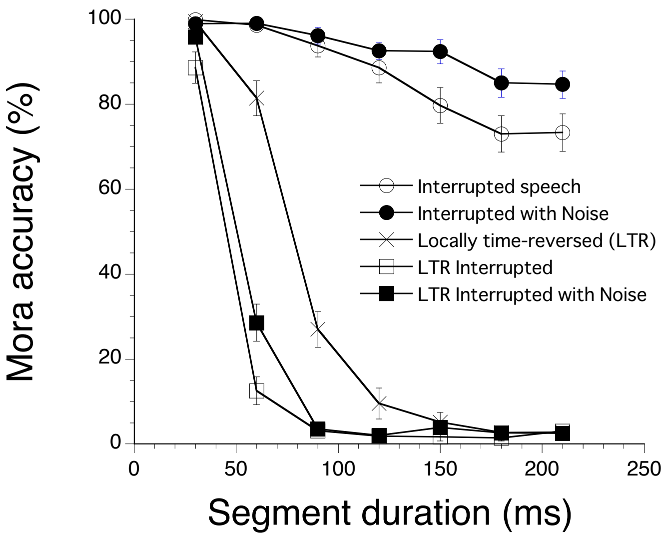


Fig. 1. Preliminary data of mora accuracy. Both normal speech interrupted with noise and locally time-reversed speech interrupted with noise showed higher mora accuracy than their counterparts that were interrupted with silence at some points of segment duration.

INFLUENCE OF THE TEMPORAL-UNIT DURATION ON THE INTELLIGIBILITY OF MOSAIC SPEECH: A COMPARISON BETWEEN JAPANESE AND ENGLISH

Kaori Kojima[1], Yoshitaka Nakajima[2], Kazuo Ueda[2], Gerard B. Remijn[2], Mark A. Elliott[3], and Sophia Arndt[3]
[1]*Graduate School of Design, Human Science Course, Kyushu University*
[2]*Department of Human Science/Research Center for Applied Perceptual Science, Kyushu University*
[3]*School of Psychology, National University of Ireland, Galway*

Intelligibility of English mosaic speech was measured in an auditory experiment. The mosaic speech was generated by dividing spoken sentences into narrow frequency bands simulating the auditory periphery and segmenting the obtained power fluctuations into equal temporal units of 20–320 ms. The average energy density within each time-frequency block replaced the original energy distribution in the block. In this study, 50% of the linguistic contents of original speech were estimated to be understood when the temporal-unit duration was about 60 ms. A significant difference between the results of this experiment and a previous experiment with Japanese mosaic speech was found. A finer temporal resolution was necessary for the English mosaic speech than for the Japanese mosaic speech. The difference between these studies seems to be caused by the difference in the speech speed, which is currently subject of further investigation.

THE FFECT OF VIRTUAL EYES ON PSYCHOLOGICAL NERVOUSNESS ON PUBLIC SPEAKERS

Miharu Fuyuno[1], Yuko Yamashita[2], and Yoshitaka Nakajima[1]
[1]*Research Center for Applied Perceptual Science and Faculty of Design, Kyushu University, 4-9-1, Shiobaru, Minami-ku, Fukuoka City, Fukuoka, Japan*
[2]*Department of English Communication, College of Engineering, Shibaura Institute of Technology, 307 Fukasaku, Minuma-ku, Saitama City, Saitama, Japan*
<m-fuyuno@design.kyushu-u.ac.jp, yama-y@shibaura-it.ac.jp
nakajima@design.kyushu-u.ac.jp>

Abstract

This study examines the nervousness of Japanese speakers when delivering public speeches in English in order to develop an effective training application for managing speakers' nervousness. Previous studies have pointed out that human behaviors tend to be affected by the presence of others because of the reputation management process. However, the presence of real people is not an indispensable factor for this effect: illustrated eye objects have also been reported to produce similar effects on humans. This research examined the effect of virtual eyes on public speakers. Participants performed short speeches in a virtual venue; questionnaire sheets were then used to measure the subjective nervousness of participants, comparing conditions with and without virtual eyes. The results suggested that virtual eyes caused significantly greater nervousness in speakers.

Humans are known to be affected by the presence of others, both psychologically and behaviorally. This fact is considered to be related to our reputation management process (Haley & Fessler, 2005). We can observe a common example of this issue in the fear of public speaking. Speaking in public is a major cause of social fear for most people, regardless of which generation they are from (Kessler et al., 1998; Pertaub et al., 2001). In modern society, skills for performing effective public speaking can influence important junctures of our lives. Effective public speaking skills are usually acquired by training and experience, and the psychological fear of speaking in front of others is a particularly difficult factor to overcome. Consequently, various studies have aimed at reducing speakers' nervousness during public speaking.

Previous studies have discussed the effectiveness of rehearsing, especially in a virtual environment where a speaker can feel nervous in a simulation of real performance. For example, Brinkman et al. (2012; 2014) presented a virtual world to participants with/without a virtual human avatar. Their experiment results showed that the presence of human avatars significantly increased participants' Subjective Units of Discomfort scores. Slater et al. (1999) used a virtual speech venue with a 3D computer graphics (CG) human audience to measure speakers' anxiety. They used three types of 3D CG audience: an audience with a neutral attitude toward the speaker, an audience with a positive attitude, and an audience with a negative attitude. Their result showed that participants felt significantly stronger anxiety with a virtual audience that had a negative attitude.

For the purpose of speech training, Fuyuno and Yamada (2016) compared a virtual audience of 3D CG, a virtual audience of a live-action movie, and a control condition without a virtual audience in terms of their effects in causing nervousness to speakers. The experiment's results showed that the virtual audience of the live-action movie caused significantly greater nervousness than the control condition. Comments from the participants suggested that both 3D CG and live-action movie audiences would be useful in providing training according to speakers' experience and proficiency.

As these previous studies indicate, the virtual presence of others can cause psychological stress for speakers, and this effect is, to some extent, useful for training in public speaking. An interesting fact here is that the degree of deformation of the human avatars was different in the above studies. For example, the CG avatars in Slater et al.'s (1999) study had relatively rough polygonal bodies, whereas Brinkman et al. (2014) used more elaborate human CG avatars. However, their results showed that the avatars' effects on humans were consistent, indicating that speakers are affected as long as they consider the avatars to be human-like objects.

In fact, a study by Haley and Fessler (2005) suggests that even a merely symbolized set of eye objects can affect human behaviors. This study involved an experiment using an economic game: Participants were divided into experiment and control groups, both of which used a computer to process the game. The experiment group was provided with a computer screen that showed a symbolized set of eye objects, whereas the control group was not shown the eye objects. Participants in the experiment group showed a significantly higher rate of generous behavior toward others in the economic game. Thus, the eye objects seemed to affect participants' behavior. In addition, Rigdon et al. (2009) examined the effect of viewing a schematic facial object consisting of three simple dots that symbolized an abstract human face watching the participants. Their study's results also suggested that the participants were affected by a subtle social cue from the schematic facial object. Considering this research background, the present study aims to examine the effects of virtual eyes on the psychological nervousness of speakers in a virtual public speech setting.

Method

An experiment was performed with 24 participants (aged 19–22 years; 17 females and 7 males). They were all non-English major undergraduate students from a Japanese university. Each participant was presented with a public speech setting of a projected virtual venue. There were three types of virtual venues: A) a venue without any eyes (only a shot of the venue with desks and chairs), B) a venue with simple line-art CG virtual eyes, and C) a venue with detailed CG virtual eyes. Figure 1 shows screenshots

(A) (B) (C)

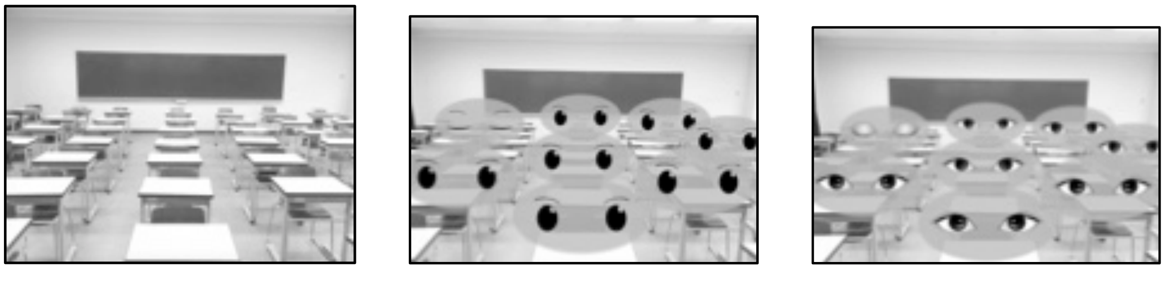


Fig. 1. Screenshots of the three types of virtual venues: A) only a venue, B) a venue with simple line-art CG virtual eyes, and C) a venue with detailed CG virtual eyes.

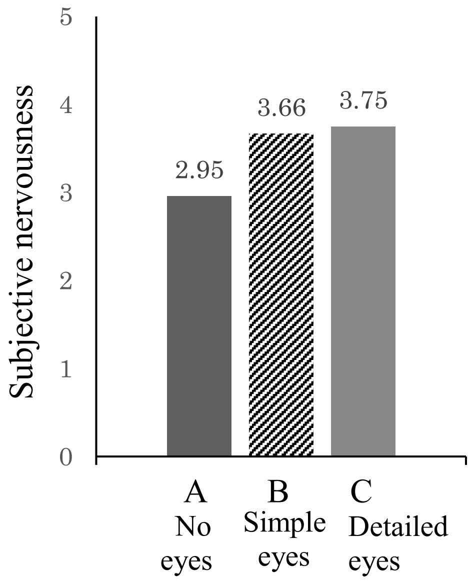


Fig. 2. Average result of subjective nervousness in each venue type.

of the three types of venues.

In both B and C, translucent oval figures were placed behind each set of eyes to emphasize the eyes against the background. The eyes were programmed to blink at a natural speed. The order of presenting the virtual venues was random for each participant.

The participants were asked to deliver speeches in English in front of the virtual venue after a preparation time of 30 seconds for each venue type. The theme of the speech was set by the experimenter. There were three kinds of themes, all regarding self-introduction: My Family, My Friends, and My Hobby. The order of the themes was random for each participant. Themes were changed for each speech in order to prevent a practice effect, which could result in a reduction in speakers' nervousness during the second and third speeches. The results of a pilot experiment confirmed that the themes were at a similar difficulty level for participants. After each speech, the participants answered a questionnaire sheet comprising a five-point Likert scale for subjective nervousness (1: I didn't feel nervous. / 5: I felt nervous.).

Results and Discussion

Result of Subjective Nervousness

The response to the Likert scale was digitized and averaged (Ms = 2.95, 3.66, and 3.75; SDs = 1.19, 0.96, and 0.98). Statistical analysis was performed with the Friedman test. The test results showed a significant difference among the three conditions ($p = .010$). The virtual venue with elaborate virtual eyes was confirmed to have caused higher nervousness than the other types. Also, the virtual venue with simple line-art virtual eyes caused higher nervousness than that without any eyes. Figure 2 graphically presents the results.

Discussion

Our finding was essentially consistent with that of the psychological experiments by Haley and Fessler (2005) in terms of the effects of eye objects on participants. The results of the present study suggested that there may be a gradation of effects according to the degree to which the objects resemble real human eyes.

In previous studies, most virtual audience characters used in speech training settings were avatars that had human faces and bodies. However, the present study's results indicated that speakers feel nervous even when they encounter virtual eyes without proper faces or bodies. Applying virtual eyes in virtual public-speaking training will be useful in increasing the applicability of the training contents, given that eye objects are more easily plotted in scenes than fully human avatars in terms of object space and data size.

Acknowledgments

This research was sponsored by the JSPS KAKENHI Grant Numbers 15K12416 and 16H03079 and grants from the Kyushu University university-initiated venture business seed development program (GAP Fund).

References

Brinkman, W., Emmelkamp, P.G., Hartanto, D., Kang, N., Kampmann, I.L., Morina, N., Neerincx, M.A., & Vliegher, D.D. (2012). *A virtual reality dialogue system for the treatment of social phobia.* CHI Extended Abstracts.

Brinkman, W., Emmelkamp, P.G., Hartanto, D., Kampmann, I.L., Morina, N., & Neerincx, M.A. (2014). Controlling Social Stress in Virtual Reality Environments. *PloS one.*

Fuyuno, M., & Yamada, Y. (2016). Development and evaluation of virtual audience movies for public speaking training. *Proceedings of 23rd Annual Meeting of Japan Association for Educational Media Study*, 30-31.

Haley, K. J., & Fessler, D.M. (2005). Nobody's watching?: Subtle cues affect generosity in an anonymous economic game. *Evolution and Human behavior*, 26(3), 245-256.

Kessler, R. C., Stein, M. B., & Berglund, P. (1998). Social phobia subtypes in the National Comorbidity Survey. *American Journal of Psychiatry*, 155, 613-619.

Pertaub, D. P., Slater, M., & Barker, C. (2001). An experiment on fear of public speaking in virtual reality. *Studies in Health Technology and Informatics*, 372-378.

Rigdon, M., Ishii, K., Watabe, M., & Kitayama, S. (2009). Minimal social cues in the dictator game. *Journal of Economic Psychology*, 30(3), 358-367.

Slater, M., Pertaub, D. P., & Steed, A. (1999). Public speaking in virtual reality: Facing an audience of avatars. *IEEE Computer Graphics and Applications*, 19(2), 6-9.

CRITICAL THINKING IN TWO LANGUAGES: BILINGUALISM AND SECOND LANGUAGE COMPETENCY AS DETERMINANTS OF PERFORMANCE

Sophia Arndt[1], Ethan Keeney[1], Kaori Kojima[2], Sarah Volkmer[3], Zhimin Bao[2], Shimeng Liu[2], Stanislava Antonijevic[1], Yoshitaka Nakajima[4], and Mark A. Elliott[1]

[1]*School of Psychology, National University of Ireland, Galway, Republic of Ireland*

[2]*Graduate School of Design, Human Science Course, Kyushu University, Fukuoka, Japan*

[3]*Faculty of Psychology and Educational Sciences, Ludwig Maximilian University Munich Germany*

[4]*Department of Human Science/Research Center for Applied Perceptual Science Faculty of Design, Kyushu University, Fukuoka, Japan*

Abstract

This study aims to examine the relationship between bilingualism and critical thinking. Monolingual English and bilingual English-Irish speakers completed an online questionnaire which assessed their ability in five dimensions of critical thinking; hypotheses testing, verbal reasoning, judging probability, argumentation analysis and problem solving. Results of this study show no significant differences between monolingual and bilingual individuals. Further analysis examined the relationship between these critical thinking measures and a number of covariates, including scores on the abbreviated Raven's Standard Progressive Matrices (RSPM) and participants' education levels. Scores on the RSPM and years of education were both found to be significant covariates.

More than half of the world's population are bilingual (Hoffman, 2014). Bilingualism is most prevalent in Europe, where 54% of all Europeans report being able to speak a second language (Eurobarometer, 2006). In an increasingly globalized world, examining for potential cognitive differences between the monolingual and bilingual population is becoming more and more relevant. Findings in this area have an extensive reach, effecting how we view the process of learning further languages and the emphasis our education system places on bilingualism.

Early research in to bilingualism generally concluded that the consequences of children growing up bilingual were negative and that learning two (or more) languages would confuse children and hinder development (Bialystok, Craik, & Luk, 2012). Peal and Lambert (1962) published a landmark study on bilingualisms effect on cognitive functioning in the early-1960's, which marked a turning point for public opinion on bilingualism. They examined how children in Montreal who were either French-speaking monolinguals or English-French bilinguals performed on a battery of tests, concluding that in the case of bilinguals, "intellectually his experience with two language systems seems to have left him with a mental flexibility, a superiority in concept formation, a more diversified set of mental abilities." (Peal & Lambert, 1962, p20).

It is now known that monolinguals and bilinguals differ in the development, efficiency and decline of fundamental cognitive abilities (Bialystok et al., 2012). Research has generally supported a favourable view on the cognitive benefits of bilingualism, however bilinguals have been repeatedly shown to have poorer vocabularies than monolinguals (Bialystok, Luk, Peets, & Yang, 2010). Studies have reported that bilinguals are more proficient in divergent thinking, creativity and flexibility of thought than their monolingual peers (Lee & Kim, 2011).

Bilingualism and Critical Thinking

An area of cognitive functioning which has received less attention in the bilingual research literature is critical thinking (CT). Critical thinking has been conceptualized in various ways over the last century. Halpern's (1998, p.450) concise definition of critical thinking, which serves as the theoretical basis for this study, describes the process as "the use of those cognitive skills or strategies that increase the probability of a desirable outcome". McPeck (1981, p.8) states that critical thinking is "the propensity and skill to engage in an activity with reflective scepticism". As a whole, definitions of critical thinking tend to share many overlapping assumptions. Researchers agree that critical thinking is a higher-order cognitive process which involves the analysis and evaluation of evidence and arguments in order to make good, well-informed decisions and solve problems. It requires a type of non-automatic, reflective responding which implies an underlying engagement of executive functioning (Noone, Bunting & Hogan, 2016). Few studies have employed a general measure of critical thinking to examine any potential difference between monolinguals and bilinguals in CT ability. A study by Albert, Albert and Radsma (2002) on bilingual and monolingual nursing students failed to provide sufficient evidence to support the existence of a relationship between bilingualism and critical thinking ability. However, Merrikhi (2011) found that in a sample of Iranian students, bilinguals were significantly more adept at critical thinking than their monolingual peers.

The Present Research

Some research has explored bilingualism's relationship with various cognitive abilities and has shown significant interactions in faculties such as in attentional control, divergent thinking and metalinguistic awareness (Cummins, 1978). Building on this research, it is a logical assumption that learning a second language may affect one's ability for critical thinking, as it has for the several other cognitive processes mentioned above. The present study aims to expand on research exploring how bilingualism affects cognition by examining whether English-Irish bilinguals differ from monolingual English-only individuals in measures of critical thinking.

This study will explore two research questions.

1) Do bilinguals score differently compared to monolinguals in measures of critical thinking?

2) Does the degree of bilingualism (language competency of L2) influence the outcome of critical thinking scores?

Method

Research Design

This study used a cross-sectional, between-subjects, correlational design to ascertain the relationship, if any, between bilingualism and critical thinking. Participants were split in to two groups based on whether they were monolingual (English only) or bilingual (English and Irish).

Participants

A convenience and snow-ball sampling strategy was used to recruit a total of 115 participants. Nine participants were excluded from the sample due to their first language being neither English nor Irish, leading to a total of 106 participants. ($N = 106$). Participants consisted of a near equal number of bilinguals ($N = 58$) and monolinguals ($N = 48$). They were accessed through three primary mediums: the National University of Ireland, Galway (NUIG), social media and Irish speaking organisations. Aside from participants' first language to be English or Irish, there were no exclusion criteria for participants in this study. Full ethical approval was granted for this study by the ethics board of the National University of Ireland, Galway. Demographic data was collected including age, gender, education, country of origin and any pre-existing medical conditions which may impact performance.

The sample was comprised of 66% females ($N = 70$), 33% males ($N = 35$), and 1% of respondents didn't identify with either ($N = 1$). The mean age was 26 ($M = 26.0$), with a standard deviation of 9 years ($SD = 9.1$). There was a near equal number of bilingual ($N = 58$) and monolingual ($N = 48$) respondents. Participants had a mean number of 16 years of formal education ($M = 16.2, SD = 3.8$). 2% ($N = 2$) of participants reported a learning disability (dyslexia), none reported a language or hearing disability, 26% ($N = 28$) reported minor vision problems.

Measures

An abbreviated nine-item form of the Raven's Standard Progressive Matrices test (RSPM; Bilker, Hansen, Brensinger, Richard, Gur & Gur, 2012) was used as a control measure for participants' cognitive abilities. The RSPM is a non-verbal measure of fluid intelligence. It uses visual geometric puzzles, each with a missing component. Participants are asked to choose between a number of options to fulfil the sequence provided. Bilker et al. (2012) report a correlation of .98 (r = .98) between the abbreviated 9- item form and it's 60-item precursor, indicating the abbreviated form predicts the original with high accuracy.

The Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld & Kaushanskaya, 2007) is a self-report language measure that incorporates both language history and proficiency. It's questions are based on factors which have been identified as important contributors to bilingual status, including "language competence (including proficiency, dominance, and preference ratings); age of language acquisition; modes of language acquisition; prior language exposure; and current language use" (Marian et al., 2007, p.943). The LEAP-Q asks participants to rate their language ability in three areas: reading, understanding and speaking on a scale of 1 to 10, were 1 indicates a minimal ability and 10 indicates strong proficiency. Marian et al. (2007) established strong internal validity in bilingual and multilingual samples across a diverse set of languages. Furthermore, they found that the LEAP-Q is a reliable predictor of performance on external, standardised measures of language proficiency across various domains (phonology, vocabulary, and morphosyntax), despite the entirely self-estimated nature of the measure.

In order to measure critical thinking ability, 'Mixed Measures' of published tests were compiled based on Halpern's (1998) theory of critical thinking. Halpern's (1998) theory discusses five dimensions of critical thinking; verbal reasoning (VR), argument analysis (AA), thinking as hypothesis testing (HT), likelihood and uncertainty (JL), and decision making and problem solving (PS). Using these dimensions of critical thinking as a basis, a search through published literature was undertaken to identify suitable assessments. A collection of tests was then developed and translated from English to Irish by a paid professional to create a mixed-measures critical-thinking test battery.

Procedure

Participants received a link which allowed them to access an online questionnaire. Before completing the questionnaire, participants were subject to an initial consent form. This consent form described the purpose of the study and full disclosure was given on the area of research. Participants' anonymity was guaranteed. Participants were assured that they were under no obligation to complete the questionnaire and were free to terminate their participation at any time. In order to progress, participants were required to indicate understanding and consent.

Both monolinguals and bilinguals were asked about demographic and socio-economic variables. Initially, participants indicated whether or not they could speak Irish. If they responded no, they would continue on to complete an English version of the critical thinking test battery. If they responded yes, then they would be subject to the LEAP-Q. The questionnaire was coded so that only those who rated their Irish (L2) fluency above 5 (on a scale of 1-10, with ten being complete fluent), would be considered bilingual for the purpose of this research. Those who met this criterion completed two versions of the critical thinking test battery and assignment of which language came first was done at random. The questions between versions were in essence the same but with slight changes to avoid repetition of answers based on previous knowledge. Distinctive features of questions such as names and defining characteristics were changed, for example if the question asked about buying a new car in the first version, it would be about buying a new computer in the second. Before submitting their data, participants were asked whether any of the questions seemed familiar. This acted as a way of knowing if bilingual individuals noticed the repetition of questions and if any of the participants had previous knowledge of the questions.

Statistical Analysis

On completion of data collection all data underwent coding and scoring on SPSS (version 23). Descriptive analysis was conducted following the collection of all completed questionnaires. All critical thinking data was normalised by using the square root of the summed score for calculations. General linear models (GLM) were conducted to examine the potential relationship between bilingualism and critical thinking. The independent variable (IV) examined in analysis was language (bilingual or monolingual). The dependent variable (DV) examined was the total overall score (Etotal) analysed through a univariate measure, and scores on the five dimensions of critical thinking analysed in a multivariate method.

Results

Descriptive Statistics

Of those who completed the bilingual section of this study, the mean self-reported fluency was 7 ($M = 7$) on a scale from 1 (none) to 10 (fluent), with a standard deviation of 2.3 ($SD = 2.3$). From the bilingual sample, 37 (35%) of respondents commented that they noticed, to varying extents, some form of repetition between assessments presented in the two languages (while the surface structure of the problems was changed the deep structure remained the same), 69 (65%) did not. Bilingual individuals ($N = 58$) were seen to score higher in the abbreviated RSPM ($M = 6.28, SD = 2.48$) than monolinguals ($N = 48; M = 5.90, SD = 2.59$). Monolingual ($N = 48$) participants mean score on the English critical thinking test battery was 20.23 ($M = 20.23, SD = 4.84$). Bilingual ($N = 58$) participants mean score was 19.78 ($M = 19.78, SD = 4.78$).

Inferential Statistics

An initial Pearson's correlation was conducted to determine if there was a correlation between individual RSPM scores (ER), which served as a control measure, and performance on the English version of the critical thinking test battery (Etotal). Scores on the RSPM were positively correlated with Etotal, significant at the two-tailed significance level ($r = .49, p < .001$).

A multiple linear regression was calculated to predict Etotal based on Age, Gender, level of education of the participant (Edu) and of the participants' parents, third language skills, recognition of question repetition in English and Irish, Irish fluency grouped into four categories and overall scores of the RSPM (ER). A significant regression equation was found for independent variables Edu and ER [$F(4.51), p < .001$; with an R^2 of .37] but not for the degree of bilingualism (language competency of L2). Both Edu ($B = .05, p < .005$) and ER ($B = .20, p < .05$) were significant predictors of Etotal. This led to the inclusion of those factors as covariates in the following calculations.

A one-way ANCOVA was conducted to determine a statistically significant difference between mono- and bilinguals on Etotal controlling for participants' education and overall scores on the abbreviated RSPM (ER). There is no significant effect of mono- or bilingualism on Etotal after controlling for Edu and ER [$F(1, 106) = .18, p > .05$]. A MANCOVA was conducted to determine a statistically significant difference between mono- and bilinguals on the sub scores of CT (HT, VR, AA, JL, PS) controlling for participants' education and ER. Wilks' lambda was significant for the two covariates Edu ($\Lambda = .83, p < .005$) and ER ($\Lambda = .77, p < .001$) but not for the independent variable MonoBilingual ($\Lambda = .97, p > .05$), indicating the absence of a significant difference between the two groups within all subcategories.

Discussion

The primary research question of this study explored whether or not bilinguals and monolinguals differ in measures of critical thinking. This study's general finding is that scores of bilinguals didn't significantly vary from monolingual scores. The degree of bilingualism (language competency of L2) did not affect these outcomes. This result was seen in both groups of bilinguals and remained constant when education of the participant and individual scoring in the RSPM were examined as potential covariates.

Years of formal education were found to have a significant influence on critical thinking scores in the monolingual and the bilingual group. They acted as a covariate for the overall CT score and the subcategories of CT in monolinguals and across both of the bilinguals' languages. Education was seen to have a positive interaction with CT scores, indicating that the bilinguals' level of education may be determinative of the individuals' critical thinking ability. This supports prior research findings including Abrami, Bernard, Borokhovski et al. (2008) who have shown the positive effect of education on critical thinking skills. The analysis of RSPM acted as a cognitive control in this study and found, supportive of the outcomes of this study, that higher scores in the Raven predicted higher scores on individual measures of the CT test battery. This was found throughout both groups, monolinguals and bilinguals.

References

Abrami, P., Bernard, R., Borokhovski, E., Wade, A., Surkes, M., Tamim, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research*, 78(4), 1102–1134.

Albert, R., Albert, R., & Radsma, J. (2002). Relationships among bilingualism, critical thinking ability, and critical thinking disposition. *Journal of Professional Nursing*, 18(4), 220–229.

Bialystok, E., Craik, F., & Luk, G. (2012). Bilingualism: consequences for mind and brain. *Trends in Cognitive Sciences*, 16(4), 240–250.

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, 13(04), 525–531.

Bilker, W., Hansen, J., Brensinger, C., Richard, J., Gur, R., & Gur, R. (2012). Development of Abbreviated Nine-Item Forms of the Raven's Standard Progressive Matrices Test. *Assessment*, 19(3), 354–369.

Cummins, J. (1978). Bilingualism and the development of metalinguistic awareness. *Journal of Cross-Cultural Psychology*, 9, 131–149.

Eurobarometer, S. (2006). *Europeans and their Languages.* European Commission.

Halpern, D. (1998). Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *American Psychologist*, 53(4), 449– 455.

Hoffmann, C. (2014). *Introduction to bilingualism.* London: Routledge.

Lee, H., & Kim, K. (2011). Can speaking more languages enhance your creativity? Relationship between bilingualism and creative potential among Korean American students with multicultural link. *Personality and Individual Differences*, 50(8), 1186– 1190.

Marian, V., Blumenfeld, H., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech Language and Hearing Research*, 50(4), 940.

Merrikhi, P. (2011). The Effect of "Bilingualism" on Iranian ELT Student's "Critical thinking ability"(CT). *Theory & Practice in Language Studies*, 1(10).

McPeck, J. (1981). *Critical Thinking and Education.* New York City: St. Martin's Press.

Noone, C., Bunting, B., & Hogan, M. (2016). Does Mindfulness Enhance Critical Thinking? Evidence for the Mediating Effects of Executive Functioning in the Relationship between Mindfulness and Critical Thinking. *Frontiers in Psychology*, 6.

Peal, E., & Lambert, W. (1962). The relation of bilingualism to intelligence. *Psychological Monographs: General and Applied*, 76(27), 1–23.

UNIQUENESS OF THE GAZE CUE IN SPATIAL ANISOTROPY OF VISUAL ATTENTION

Hirokazu Eito
Graduate School of Advanced Integration Science, Chiba University
1-33 Yayoi-cho, Inage-ku, Chiba, Japan

Akio Wakabayashi
Graduate School of Humanities, Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba, Japan

We investigated if spatial attention was changed to location-based or object-based by the type of cues, social or non-social. Experiment 1, participants shift their attention to one end of one of two rectangles by central cues in arrow, eye gaze, or pointing with finger, just before they distinguished the target presented peripherally. We compared the response times to check if attention was orienting to the cued location or the cued object by each cue type. Results showed that only gaze cues did not cause object-based attention, while we found an anisotropy of attention only in the gaze cue condition. This anisotropy was found in all cue types when object was removed in Experiment 2. These results suggested that the anisotropy of attention by the gaze maintained strongly in the context of object-based attentional selection, and that results, which is unique to the gaze, affected to the type of attentional selection.

MOMENTARILY VIEWING AN AVERTED GAZE SHIFTS VISUAL ATTENTION

Masaki Ogawa[1], Hiroyuki Ito[1,2], and Shoji Sunaga[1,2]
[1]*Faculty of Design, Kyushu University, Fukuoka, Japan*
[2]*Research Center for Applied Perceptual Science, Kyushu University, Fukuoka, Japan*

We confirmed that viewing an averted gaze for only a very brief moment can induce an attentional shift. We used face stimuli from four individuals that displayed either direct or averted gazes. A direct-gaze stimulus was presented for 500–1000 ms, followed by an averted-gaze stimulus for 16.7, 33.3, or 66.7 ms, and then a return to the direct-gaze stimulus. Participants were asked to indicate the direction (left or right) of a target stimulus by pressing a button as quickly as possible. Targets were displayed 0, 200, or 1000 ms after the averted-gaze stimulus. The results showed all three averted-gaze conditions induced the same degree of attentional shifts, as evidenced by reaction time to the target. However, when participants were asked to indicate the direction of the averted gaze, performance in the 16.6 ms averted-gaze condition was lower than the 33.3 and 66.7 ms conditions. These results may indicate that subliminal perception of gaze direction shifts visual attention.

INTEROCULAR INDEPENDENCE OF ATTENTION IN PIGEONS

Aya Kokubu and Kazuo Fujita
Graduate School of Letters, Kyoto University
<kokubu.aya.77w@st.kyoto-u.ac.jp>

Pigeons have laterally positioned eyes and 2 foveas in each eye. These features suggest that pigeons have evolved to see four positions: left front, right front, left lateral and right lateral. An interesting question is whether they perceive different visual scenes simultaneously. Kaneko (2013) examined this question by training pigeons on a Posner task, which presented a cue and a target either ipsilaterally or contralaterally within or across visual hemifield. The result showed that cueing effect to RT was restricted within the same hemifield, suggesting that pigeons have independent attentional resource for each hemifield. We examined whether pigeons have independent attentional resources for each eye, in frontal binocular field. We separated visual input onto both eyes by using anaglyph glasses and conducted a Posner task. If pigeons have independent attentional resource for each eye, cueing effect may be restricted within one eye. This is an on-going experiment, so the result will be presented at the meeting.

FACTORS AFFECTING THE PERIPHERAL FLICKER ILLUSION

Hiroyuki Ito[1,2] and Tomomi Koizumi[3]
[1] *Faculty of Design, Kyushu University, Fukuoka, Japan*
[2] *Research Center for Applied Perceptual Science, Kyushu University, Fukuoka, Japan*
[3] *Graduate School of Design, Kyushu University, Fukuoka, Japan*

We have discovered a new illusion—termed the Peripheral Flicker Illusion—and investigated factors that affect the phenomenon. When a green or blue object (2.0 cd/m^2) was presented on a red background (6.0 cd/m^2), the object was seen to flicker. Red objects never caused this effect no matter what the background color, nor did green or blue backgrounds no matter what the object color. The flickering illusion increased with the object's degree of retinal eccentricity (max 20°), with hardly any effect when it was presented centrally. The luminance of the object was also an important factor. The best luminance for the green and blue objects was 3.1 and 1.2 cd/m^2, respectively. We think that the apparent flicker is caused by delayed responses of rods interacting with L-cones.
(Supported by the JSPS)

VISUAL SYSTEM RESITS VERTICAL HORIZONTAL ILLUSION AND PONZO ILLUSION BY MATHEMATIC OBSERVATION

Masahiro Ishii
School of Design, Sapporo City University, Sapporo, Hokkaido, Japan
<m.ishii@scu.ac.jp>

Aglioti et al. reported that a geometrical-optical illusion had no effect on the preshaping aperture. Franz et al., however, reported that grip scaling was sensitive to the illusion. This discrepancy leads to an assumption that the visual system has two representations, and the visuomotor system uses one of them in accordance with the task. In the current study, I investigate if the visual perception system behaves in the same fashion. When a figure is presented, we usually judge the line length from its holistic appearance. We also can mentally rotate and translate the bisected horizontal line to compare with the one-half of the vertical line. Experiments were conducted using the vertical-horizontal illusion and the Ponzo illusion. The results indicate that the visual system resists the illusions by mathematic observation. This suggests the visual system has two representations and the perception system uses one of them in accordance with the task.

SENTIMENT ANALYSIS ON ASSOCIATED COLORS BY LISTENING SYNTHESIZED SPEECH

Atsuya Suzuki, Win Thuzar Kyaw and Yoshinori Sagisaka
Graduate School of Fundamental Science and Engineering, Waseda University

Abstract

Aiming at cross-modal information mapping, color association characteristics have been analyzed using synthesized speech. As the previous studies on color association of natural speech has shown strong correlations between F0 and value, sound pressure level and saturation, the association between spectral features and hue values have been analyzed in this paper. To effectively understand speech spectrum-color mapping characteristics, speech synthesized with three formants and band widths were employed. Color association has been asked for 26 subjects after listening 300 speech samples with various formants and band widths. Statistical analysis has shown clear correlations between the selected colors and F1, F2 but not for others. By feeding these data to a feed-forward neural net, an associated color chart has been obtained in F1-F2 place, which visualizes speech spectrum-color association characteristics.

1. Introduction

Color association by listening sound has been analyzed using speech as listening samples for ordinary subjects not having any particular synesthesia. Word et al found that there is a general tendency to associate high pitch sounds with light colors and low pitch with dark colors[1]. Wremble et al have studied associative colors of English 12 vowels for Polish subjects whose second language is English. Through their studies, they found there is a tendency to associate light color (yellow, green, and orange) to the front tongue vowel, dark color(brown, blue, and black) for the lingual tongue vouel, and gray for the middle tongue vowel[2]. By employing HSV(Hue, Saturation, Value) description for colors instead of individual color categories, we could have found very strong correlations between F0 and value(0.90), loudness characteristics and saturation (0.85) [Watanabe et al ICPhS 2015]. For the correlations between Hue and spectrum, association tendencies have been red or orange for /a/, yellow for /i/, blue, green or purple for /u/, green or orange for /e/ and blue, green, or purple for /o/[3].

As shown above studies, very similar color association characteristics have been observed. However there is no quantitative study on the change of speech intrinsic characteristics such as vocal tract resonance (formant) or other voice characteristics and no trials of cross-modal mapping. In this paper, we have tried to specify the acoustic features relating to this spectrum-hue association using synthesized speech. By employing Formant synthesizer to substitute natural speech, we simplified control parameters, three formants and their bandwidths to analyze their correlations with hue values.

In the following Section 2, we describe the speech-color association experiment. First, we confirm the validity of the following experiments using synthesized speech samples in Section 2.1. Here we compere the color association results between natural speech and the synthesized speech. In the following Section 2.2, we explain the synthesized speech samples we employed in in this experiment. In Section 2.3, we explain our color association experiment. In Section 2.4 we confirm the validity of the conversion from Hue to RGB and describe how to do it. In the following Section 3, we show the result of experiment in terms of

Table 1
(natural speech's value / synthesized speech's value) in each vowels

	/a/	/i/	/u/	/e/	/o/
H	1.09	0.866	0.651	1.19	0.9
S	0.903	0.811	0.749	0.986	0.876
V	0.778	0.900	0.858	0.988	0.830

which parameters affect to associated colors and how they affect to associated colors. Finally in Section 4, we summarize the result and show prospects of speech-color association studies.

2. Speech-color association experiment

2.1. Comparison between natural speech and the synthesized speech

Before carrying out speech-color association experiment using synthesized speech, we confirmed the validity of employing Formant synthesizer for our research purpose. By comparing associated colors between natural speech and synthesized one with the same formants, we measured the differences of them. Arbitrary chosen 10 samples male speech consisting of five Japanese vowels (/a/, /i/, /u/, /e/, and /o/) were employed for the experiment. Three formants, bandwidths and F0 have been extracted from these samples and employed to drive the formant synthesizer. All loudness level were adjusted to around -12dB and F0 to 150Hz to eliminate their contributions to color selection.

The comparison of HSV values between the natural speech and the synthesized one in shown in Table 1. As the Table shown, each values are close to 1, so we could not find any particular differences between natural speech and synthesized speech. Though the experiment size is not so big, from this result, we think it reasonable to employ synthesized speech for our following speech-color association experiments.

2.2. Color association experiment using the synthesized speech

For voice stimuli, we synthesized 300 speech samples by randomly selecting F1 to F3 frequencies and their Q values. This random selection was carried out employing uniform random number on Mel frequency scale. This is because increasing entropy on Mel frequency scale makes them in perceptually effective manner. As known from this formant parameter selection, resultant speech samples may not correspond to any realistically existing vocal tract resonance characteristics.

The Q value is obtained by dividing the bandwidth of the formant by the center frequency of it. Q is a value indicating the sharpness of a formant. We judged that the Q value does not have a marked auditory effect like a formant frequency and can not be expected to have a significant influence on associative color due to continuous change. So We made the Q value to be chosen at random from among the three classes 20, 25, 30. F0 frequency, voice source waveform and sound pressure were set as constant values. Although the strength of each formant is considered to be an important parameter which affects formant characteristics, since auditory changes were very slighter than other parameters, we set them as constant values concerning increase of dimensions of the parameters to be handled. We employed LF model as voice source[4].

2.3. Method

Subjects were 26 (13 males and 13 females) persons from 20 to 50 years old regardless of their first language. To reduce the burden on the subjects, they were divided into three groups consist of 9 subjects, 9 subjects and 8 subjects. The speech samples were divided into three groups of 100 samples. Each subject listened to 100 speech samples of the group. Therefore, 9 or 8 answers were obtained for each sample. The subjects were asked to be seated in front of the computer screen and were instructed to listen to the speech samples in random order. Then Subjects were asked to choose the most suitable color for their imagined-color on PCCS color system at a time after listening one vowel sound.

2.4. Conversion from Hue to RGB

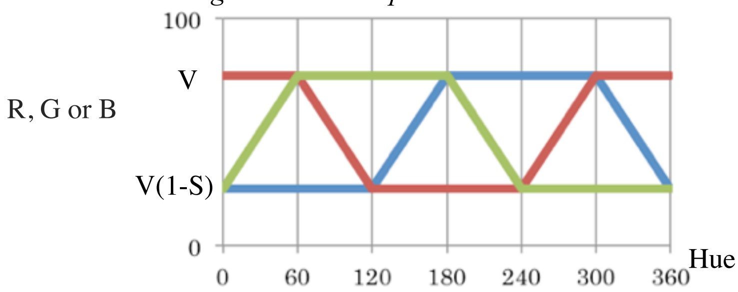
Since the Hue value represents the angle on color cycle, it sometimes cause calculative problems. In the following Section 4, the correlation between each parameter and associative color would be calculated. However, since Pearson product-moment correlation coefficient is a value indicating the magnitude of the linear relationship, values such as angles cannot be applied without adjustment. Therefore, in order to simplify the calculation, conversion of hue values to RGB values was tried.

First, the conversion from the HSV to the RGB is performed by the following equation.

$$H' = \frac{H}{60°},\ X = S(1 - |H' \bmod 2 - 1|),$$

$$(R,G,B) = (V-S)(1,1,1) + \begin{cases} (0,0,0) & (\text{if H is undifined}) \\ (S,X,0) & (\text{if } 0 \leq \text{H}' < 1) \\ (X,S,0) & (\text{if } 1 \leq \text{H}' < 2) \\ (0,S,X) & (\text{if } 2 \leq \text{H}' < 3) \\ (0,X,S) & (\text{if } 3 \leq \text{H}' < 4) \\ (X,0,S) & (\text{if } 4 \leq \text{H}' < 5) \\ (S,0,X) & (\text{if } 5 \leq \text{H}' < 6) \end{cases}$$

Fig 1 Relationship between Hue and RGB



Therefore, the behavior of increasing and decreasing of each of R, G, B depends only on Hue value. Hence the result will not be defferent even if Value and Saturation are chosen arbitrary.

3. Results

3.1. The correlation between each parameter and associative color

In order to overview how each parameter of speech affects associative color, correlation coefficient was calculated firstly. The results are shown in Table 2. In the table, *F1* to *F3* are center frequencies of F1 to F3, *Q1* to *Q3* are Q values of F1 to F3, and *Qsum* is a sum of Q1 to Q3. From these results, it can be seen that the frequencies of F1 and F2 have a stronger

table2 Correlation coefficient between each parameter of speech and RGB value

	F1	F2	F3	Q1	Q2	Q3	Qsum
R	***0.467***	0.141	0.087	-0.042	-0.004	-0.004	-0.028
G	0.152	***0.387***	0.062	0.038	0.086	0.139	0.106
B	**-0.544**	***-0.371***	-0.120	0.058	-0.057	-0.026	-0.015

correlation with associative color than other parameters. Although the other parameters have low correlation coefficients, it is possible that they may affect associative colors due to interaction or nonlinear relationship, so it cannot be concluded that they are irrelevant at this point.

Therefore, the relationship was examined by comparing the prediction accuracy using machine learning performed by the neural network model described later in this section.

Multiple regression analysis was performed employing the values of R, G and B as objective variables and F1 and F2 as explanatory variables based on the data obtained in this experiment. The results are shown in Table 3.

Table3 results of multiple regression analysis

R

	Coefficients	Standard Error	P-value
intercept	5.37×10^{-2}	5.14×10^{-2}	2.97×10^{-1}
F1	7.77×10^{-1}	8.47×10^{-2}	2×10^{-16}
F2	2.73×10^{-1}	1.02×10^{-1}	7.86×10^{-3}

Adjusted R Square：2.31×10^{-1}

G

	Coefficients	Standard Error	P-value
intercept	3.01×10^{-1}	5.03×10^{-2}	6.29×10^{-9}
F1	2.31×10^{-1}	8.28×10^{-2}	5.59×10^{-3}
F2	7.27×10^{-1}	9.97×10^{-2}	2.86×10^{-12}

Adjusted R Square：1.66×10^{-1}

B

	Coefficients	Standard Error	P-value
intercept	1.04	4.57×10^{-2}	2×10^{-16}
F1	-9.29×10^{-1}	7.54×10^{-2}	2×10^{-16}
F2	-7.55×10^{-1}	9.07×10^{-2}	3.1×10^{-15}

Adjusted R Square：4.26×10^{-1}

It can be confirmed that the P-value of each coefficient is sufficiently low. The correlation coefficient between the predicted value and the actual value of the associative color obtained from each multiple regression equation was 0.56.

In this research, a three-layer feed forward type neural network was employed. Six units corresponding to each frequency and Q value of F1, F2, F3 are placed in the input layer, and three units corresponding to the RGB value of associative color are placed in the output layer. The performance of the model was evaluated to compare the predicted value and the measured value by the cross validation method with two groups as learning data and one group as the test data. The results of the learning experiments are shown in Fig. 2. The number of units in the middle layer and the number of learning times are on the horizontal axis, correlation coefficient between the measured values and the predicted values of the associative color are on the vertical axis. The generalization result tends to worsen as the number of units in the middle layer is increased. So, we decided that two units in the middle layer is sufficient for this time. And it does not diverge even if the learning iteration is increased, but it converges to a nearly constant value, so you can see that learning is going smoothly without causing excessive learning. Therefore, this time we decided that learning iteration be 50000 times.

Fig 2 Relationship between Unit number / Learning iteration and Generalization ability

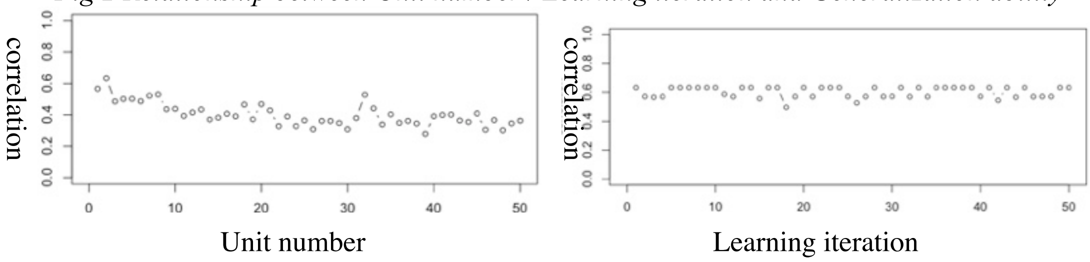
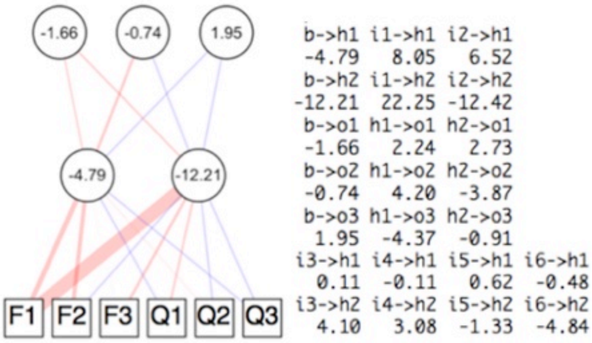


Fig 3 NeuralNetworkModel and List of weights

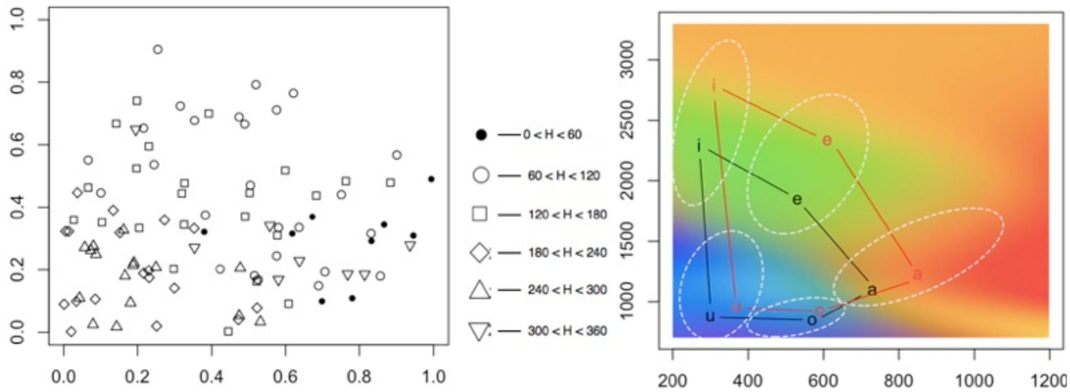


After that, the weights of each unit were adjusted to obtain the model shown in Fig 3. Each weights are also described. Learning was carried out using this model and regression prediction was performed in each of the three groups. As a result, correlation coefficients between the measured value and the predicted value were 0.63, 0.63, 0.47. From the above results, it can be seen that the weight of the unit corresponding to parameters other than the F1 and F2 frequencies in the neural network is smaller and there was no significant difference in the accuracy of multiple regression prediction performed in Section 3.4 and regression prediction using the neural network. Therefore it can be considered that those other than F1 and F2 frequencies have little effect on hue values of associative colors.

3.2. The color map

From the above it was confirmed that associative color can be predicted only by F1 and F2 frequencies, so we organized the result as a color map on F1 - F2 coordinates. This is shown in Fig 4. Moreover, by regression prediction of associative color using machine learning at arbitrary point on the coordinate, we complemented the color map. Machine learning was done employing F1, F2 frequency and RGB values of these 300 pieces of data as learning

Fig 4 Scatter plot and Color map



data. Furthermore, the average formant frequency for 5 Japanese vowel sounds of men and women is shown to this color map, witch is Fig3. From this figure, it can be seen that there might be different associative colors are more likely to be selected depending on the formant frequency difference, even if the sound is same vowel category.

4. Conclusions

In this study, to find acoustic features relating speech-color association, we employed synthesized speech with random Formants. Speech-color association experiments were carried out and the results were compared with those previously obtained using natural speech. The results are summarized as follows.

(1) The same speech-color association characteristics were observed for synthesized speech.
(2) Hue value of the associated color correlate with the F1 and F2 but not others.
(3) The color map in F1-F2 plane obtained through a neural net training quantitatively reveals speech-color association characteristics, which can give a possible speculation why multiple colors are associated with the same vowels.

As well-known the F1-F2 plane relates speech articulators (e.g. [5]), it may be possible for us to associate this correspondence as articulation-color mapping such as jaw open-close to red, orange- light, navy blue. The current results were obtained by averaging all subjects' association results. Further studies are under planning to precisely analyze them from listener's difference view point together with other sentiment correlation analyses between voice source characteristics and textures.

References

[1] J.Word, et al., "Sound - color synthesia: to what extent does it use cross- modal mechanism common to us all?", Cortex,42, pp.264-280, 2006.

[2] M.Wrembel, K. Rataj, "Sounds like a rainbow - sound-color mappings in vowel perception-", Proceedings of ISCA Tu-torial and Research Workshop on Experimental Linguistics, pp.237-240,2008.

[3] Kanako watanabe, Yoko Greenberg and Yoshinori Sagisaka " Cross-modal description of sentiment information embedded in speech " ICPhS, 2015

[4] Fant, G. and Liljencrants, J. and Lin, Q. "A four-parameter model of glottal ow", pp.001-013,1985.

[5]S.Maeda,"Compensatoryarticulationduringspeech: Evi-dencefrom the analysis and synthesis of vocal tract shapesusing an articulatory model ", Speech production and speech modeling, Kluwer Academic Publishers, pp.131-149, 1990.

A CASE STUDY ON SYNESTHESIA: IS THERE A CONNECTION BETWEEN JAPANESE HIRAGANA-KATAKANA ARTICULATION CATEGORIES AND COLOR GROUPING?

Miki Hokajo[1], Yoshitaka Nakajima[2], Kazuo Ueda[2], Chihiro Hiramatsu[2], Moe Nishikawa[3], and Gerard B. Remijn[2]

[1]*Department of Human Science, Graduate School of Design*
Kyushu University, Fukuoka, Japan

[2]*Department of Human Science/Research Center for Applied Perceptual Science*
Kyushu University, Fukuoka, Japan

[3]*Department of Communication Design Science, Graduate School of Design*
Kyushu University, Fukuoka, Japan

Grapheme-color synesthesia is a form of synesthesia in which individuals perceive a color when they see a letter, word or number (a grapheme). In studies on grapheme-color synesthesia conducted with English graphemes, where the same letter can have different pronunciations (cat, cinema), color was associated with the unique shape of a letter, not with the pronounced sound. Japanese grapheme-color synesthetes, by contrast, strongly associate colors with the pronounced sound of a ('hiragana' or 'katakana') character and not with the visual character shape. The reason is that in Japanese writing each unique syllabary sound is represented by a hiragana and a katakana character with a different visual shape (for example, さ and サ for /sa/). The present case study shows that this "phonetics-to-color" association occurred in 3 Japanese grapheme-color synesthetes even when auditory speech stimuli were presented without any visual information. That is, they perceived color just by listening to the speech sounds, which elicited a colored visual representation of the Japanese character shape in their mind's eye. Furthermore, speech sounds in similar articulation categories seemed to elicit similar color groupings, suggesting a very strong audiovisual connection in Japanese grapheme-color synesthetes.

PERSONAL SPACE DURING OBSTACLE AVOIDANCE: THE EFFECT OF AVOIDANCE DIRECTION

Kotaro Shimizu[1], Yusuke Kondo[2], Yuriko Kihara[1], Kazunari Ito[1], Keita Tai[1], Takahiro Higuchi[3], and Taketo Furuna[4]

[1] *Graduate school of Health Science, Sapporo Medical University South 1, West 17, Chuo-ku, Sapporo, Hokkaido, Japan*

[2] *Higashi Kuyakusyomae Orthopedics Hospital, 7-1-35, Kita12-johigashi Higashi-ku Sapporo-shi, Hokkaido, Japan*

[3] *Department of Health Promotion Science, Graduate School of Human Health Science Tokyo Metropolitan University, 1-1, Minamiosawa, Hachioji-shi, Tokyo, Japan*

[4] *Department of Physical Therapy, School of Health Sciences, Sapporo Medical University South 1, West 17, Chuo-ku, Sapporo, Hokkaido, Japan*

Abstract

Twelve right-handed young males participated in this study. Participants were instructed to walk normally and avoid an obstacle by passing either to the right or left of it. We used the VICON-MX system to measure obstacle avoidance and personal space distance with a 1° *spatial resolution. Personal space distance was binned into* 10° *categories based on the angle. We compared the personal space measures between right and left obstacle avoidance conditions. Results showed that lateral personal space was greater when passing to the right of obstacles, which indicates that handedness likely affects personal space distance when passing obstacles. Although measures of anterior personal space showed important individual differences, they did not differ as a function of obstacle avoidance direction.*

Collisions while walking can lead to falls and serious injuries, particularly for older individuals. To avoid collisions, individuals usually maintain a minimum distance between themselves and other individuals or obstacles (Gerin-Lajoie, 2005). This distance is often defined as "personal space", and is influenced by both the environment and by personal characteristics such as age, sex, and personality (Hayduk, 1983). Although individuals are known, in general, to often change their walking path direction and/or rotate their shoulders to maintain a minimum personal space to avoid collisions, the specific characteristics of this personal space during obstacle avoidance remain unknown. To address this knowledge gap, the present study aimed to investigate the personal space measured during obstacle avoidance as a function of avoidance direction.

Method

Twelve right-handed males (age: 21.4 ± 0.8 years, height: 170.8 ± 4.8 cm, weight: 58.2 ± 4.6 kg, shoulder width: 41.0 ± 1.7 cm) participated in this study. Written informed consent was obtained from participants prior to data collection. The study was approved by the Sapporo Medical University Ethics Committee (approval number: 28-2-26).

We set up a 9-m walking path where a vertical cylindrical pole made of polystyrene (200 cm $\times \Phi$ 30 cm) was placed 3 m from the starting point of the walking path. Obstacle avoidance was measured using the VICON-MX system (Oxford Metrics, Oxford, UK) and 10 motion capture cameras. Infrared light reflecting markers were attached to both the

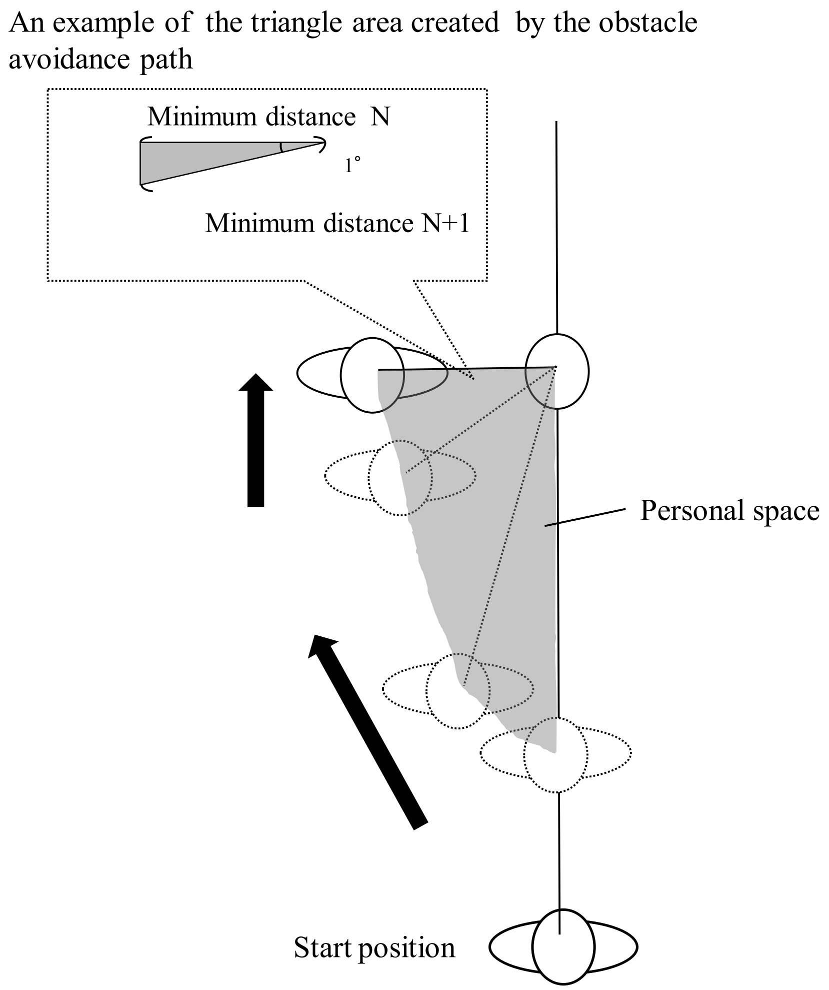


Fig. 1. The calculation of personal space during obstacle avoidance.

participant's sternum and the obstacle. We calculated the distance between a participant and an obstacle throughout the each trial. Likewise, the angle between a participant and an obstacle was measured. The baseline of angle was defined as a line drawn vertically from obstacle to walking path. Moreover, we picked out the minimum distance between a participant and an obstacle with respect to 1°, and we calculated the area of triangle consisted of two lines calculated as minimum distance. Personal space was summed up each area of triangle in each obstacle avoidance conditions (Figure 1).

The personal space was binned as a function of the following angle categories: 0°–10°, 10°–20°, 20°–30°, 30°–40°, 40°–50°, 50°–60°, 60°–70°, 70°–80°, 80°–90°. A total of two obstacle avoidance conditions were performed by each participant: one condition where they were required to pass the object to the left, and another to the right. Three trials were performed per obstacle avoidance condition. The average of personal space was compared between right and left avoidance directions using paired t-tests. A p value of less than 0.05 was considered statistically significant. All statistical analyses were performed using the SPSS statistics software version 20 (IBM Japan Ltd, Tokyo Japan).

Table 1. Comparison of personal space for left and right obstacle avoidance conditions.

Personal space	Left (m^2)	Right (m^2)	p value
0°–10°	0.0078 ± 0.0022	0.0093 ± 0.0029	0.0060*
10°–20°	0.0083 ± 0.0023	0.0098 ± 0.0032	0.016*
20°–30°	0.0094 ± 0.0027	0.011 ± 0.0036	0.043*
30°–40°	0.011 ± 0.0034	0.013 ± 0.0044	0.139
40°–50°	0.015 ± 0.0049	0.016 ± 0.0059	0.411
50°–60°	0.035 ± 0.013	0.035 ± 0.015	0.930
60°–70°	0.035 ± 0.013	0.035 ± 0.015	0.930
70°–80°	0.064 ± 0.025	0.068 ± 0.033	0.502
80°–90°	0.18 ± 0.088	0.20 ± 0.088	0.312

Results and Discussion

The average personal space measured in all conditions are presented as a function of angle category in Table 1. Paired t-tests showed significant differences for 0°–10° ($p = 0.006$), 10°–20° ($p = 0.016$), 20°–30° ($p = 0.043$).

These results indicate that participants' lateral personal space during obstacle avoidance is dependent on the chosen avoidance direction, where participants tend to maintain larger distances from the obstacle on their left (non-dominant) side, possibly due to a role played by the handedness of our participants, which is consistent with the findings of previous reports (Conson, 2010). However, the anterior personal space was not found to differ between left and right avoidance conditions. Rather, we found that the anterior personal space varied widely across participants. We believe this variance reflects perceptual and planning individual differences (Hackney, 2011), which likely play an important role in determining personal space in the anterior-posterior axis.

Although previous studies have shown that gender, closeness, and brightness are related to the dimensions of personal space (Hayduk, 1983) (Adams, 1991), they did not, however, investigate how these dimensions evolve during walking. Consequently, further studies investigating the effects during walking are needed.

Acknowledgements

The authors declare that they have no conflicts of interest. We thank Edanz Group (www.edanzediting.com/ac) for editing a draft of this manuscript.

References

Adams L., & Zuckerman D. (1991). The effect of lighting conditions on personal space requirements. *The Journal of General Psychology,* 118, 335-340.

Conson M., Aromino A. R., & Trojano L. (2010). Whose hand is this? Handedness and visual perspective modulate self/other discrimination. *Experimental Brain Research*, 206, 449-53.

Gérin-Lajoie M., Richards C. L., & McFadyen B. J. (2005). The negotiation of stationary and moving obstructions during walking: anticipatory locomotor adaptations and preservation of personal space. *Motor Control*, 9, 242-269.

Hackney A. L., & Cinelli M. E. (2011). Action strategies of older adults walking through apertures. *Gait Posture*, 33, 733-736.

Hayduk L. A., & Mainprize S. (1980). Personal space of the blind. *Social Psychology Quarterly*, 43, 216-223.

LATERAL MARGIN WHEN WALKING THROUGH AN APERTURE CREATED BY A PERSON AND A WALL

Kazunari Ito, Kotaro Shimizu, Toshiyuki Kikuchi, Yusuke Kondo, Yuriko Kihara, Keita Tai, Takahiro Higuchi, and Taketo Furuna
Graduate School of Health Sciences, Sapporo Medical University, S1 W17, Chuo-ku, Sapporo, Hokkaido, Japan
<k.itou@sapmed.ac.jp>

Many studies have investigated walking through an aperture between walls or between people. However, few studies have examined walking through apertures created between a wall and a person. This study examined the behavior of pedestrians as they walked through an aperture created between a wall and a person. The participants were 12 right-handed healthy males. The subjects walked through narrow apertures under three conditions: wall and wall (WW); wall and person (WP); and person and wall (PW). We measured the distance between the subject and the lateral obstacle (wall or person) when the subject passed through the aperture. We compared the lateral margin between the subject and lateral obstacle under the three conditions. The lateral margin was significantly different between the WP and PW conditions. Our results indicate that pedestrians tend to maintain a lateral margin when passing another person.

THE LOCATION OF VISUAL FEEDBACK INFLUENCES PERCEIVED ACTION-OUTCOME INTERVAL

Kentaro Yamamoto
Faculty of Human-Environment Studies/Research Center for Applied Perceptual Science, Kyushu University, Fukuoka, Japan

Temporal interval between an action and its sensory consequence seems to be distorted by predictability of the action outcome. Here, I examined whether the spatial location (i.e., eccentricity) of a visual consequence of action influences perceived interval between the action and outcome. Participants were asked to touch or click on a central object in the touch panel display. After the delay of 100, 400, or 700 ms, a visual feedback (rectangle) was displayed in one of four corners at an eccentricity of 4, 8, or 12 deg from the central object. Then, participants estimated the delay on a 10-point scale. I found that the delay of the visual feedback was perceived to be longer as the eccentricity became larger, independent of the approaching method (Fig. 1). The results suggested that predictability of the feedback location can modulate the perceived action-outcome interval.

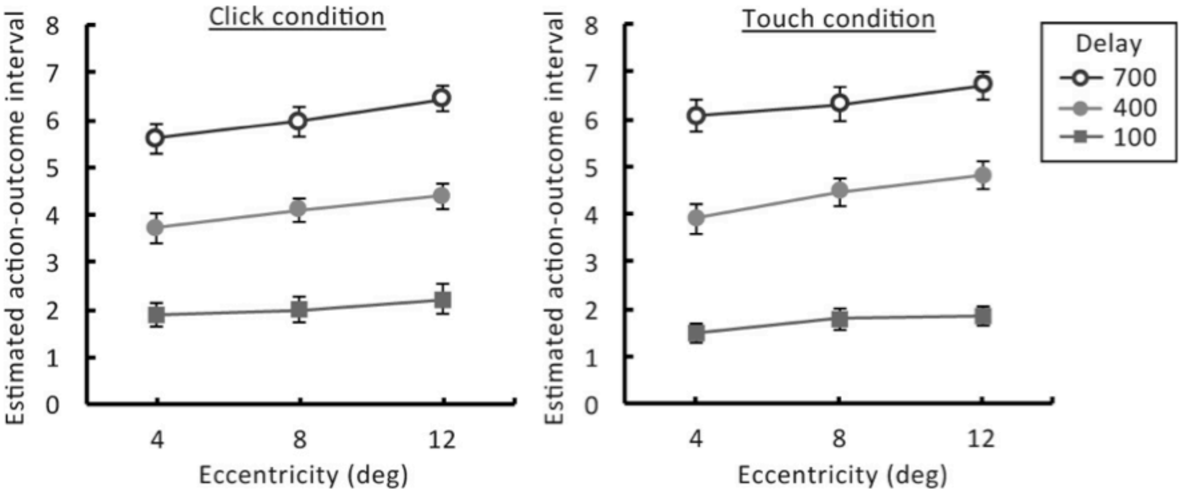


Fig. 1. The results of the present study. The estimated action-outcome interval increased with the eccentricity of visual feedback [$F(2,18) = 26.62, p < .001$] and actual feedback delay [$F(2,18) = 103.57, p < .001$].

WHERE AND WHEN DID IT BOUNCE?

Junji Yanagi and Makoto Ichikawa
Faculty of Letters, Chiba University, Japan

Abstract

When vertically aligned dots of different luminance contrasts move horizontally and abruptly reverse (bounce) the direction at the same time, they can be perceived neither aligned nor synchronized at the instant of the bounce. This perceptual phenomenon appears to be counterintuitive because both aspects of position and timing of the bounce are misperceived concurrently, rather than only one of them. In the experiment, we manipulated the spatial and temporal shift of bounces between two moving stimuli of different luminance contrasts and asked observers to make spatiotemporal judgments. Observers tended to perceive that the low-contrast stimulus bounced at an "overshoot" position and later than the high-contrast stimulus. Although this result is consistent with flash-lag effect (Kanai, Sheth, & Shimojo, 2004), further investigation will be necessary to examine whether the temporal aspect of our results can be explained by their theory.

When we see an object in motion, we perceive it at a particular position and at a particular moment. Even though we can appropriately interact with such moving objects, many researches have shown that we can often misperceive some aspects of the movement. One example of such misperception is the flash-lag illusion (Nijhawan, 1994). In this illusion, the perceived positions of moving and static stimuli appear not aligned despite that they are actually aligned when the static stimulus flashes. The footstep illusion (Anstis, 2001) is another example, where two moving stimuli of the same speed but different colors (light yellow or dark blue) appear to move in different speeds depending on whether they are on the black or white part of the striped background. This illusion can be related to the characteristic of speed perception that depends on the luminance contrast of moving stimulus (Thompson, 1982). We might as well consider the speed misperception as the mislocalization of moving stimulus. The stimuli of slower misperception can be mislocalized at the backward-shifted position from the actual position.

In this study, we tried to measure both spatial and temporal aspects of a moving object using one stimulus. If only the spatial mislocalization is the aspect to measure, we can use a spatial probe like flash-lag situation. However, this method can't assess the temporal characteristics of the moving stimuli. We can't pick a particular moment from continuous moving stimulus.

Here we use an abrupt event, especially the reversal of direction (we call it "bounce") of moving stimulus itself to investigate the both spatial and temporal aspects. The position and timing judgment of the bounce can be shifted if the ongoing process of the moving stimulus is spatially displaced or temporally delayed. Since a judgment must be done against a reference, we prepare two stimuli, target and reference, and examine the perception of spatiotemporal relationship between them as a key to elucidate the characteristics of visual motion processing.

During making demonstrations, we found that moving stimuli appear to bounce in a counterintuitive manner. We had thought that the bouncing position of a low-contrast stimulus should "undershoot" relative to a high-contrast one, since the low-contrast stimulus appears to move slower, which can be interpreted that the position should be misplaced

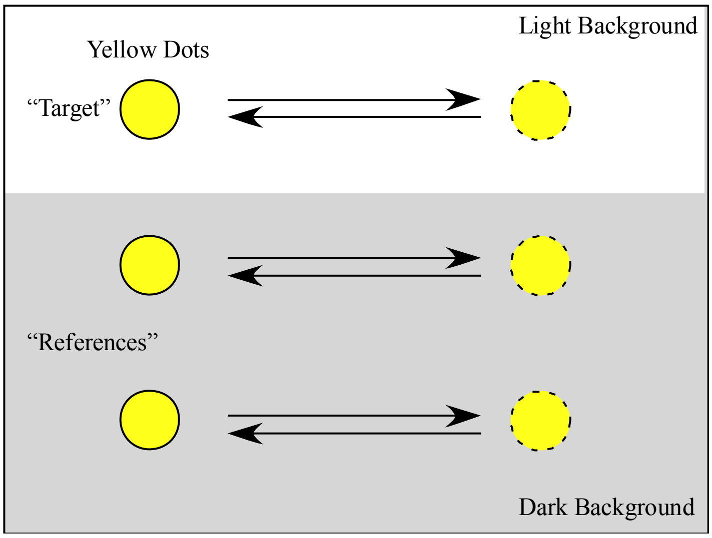


Fig. 1. Display configuration for the preliminary observation. In the actual display, the edges of dots were absent and the references' background was much darker. Observers observed the oscillating movement but judged the bouncing at the right end.

backward. However, in a preliminary observation, the bounce position of the low-contrast stimulus actually appeared to be displaced into an "overshoot" direction, going farther than the high-contrast stimulus until it bounced. So we conducted an experiment to check the consistency of this perception with a number of observers.

Observation

In order to confirm the robustness of this spatiotemporal misalignment illusion, we conducted a survey in the form of questionnaire. In a university class, we presented a movie on a large projector screen (about 2 m wide) and asked 20 students to judge the position and timing relationship of bouncing stimuli. The stimulus configuration is illustrated in Figure 1. Of the three yellow dots moving horizontally, the target was one on the light background and the references on the dark. All dots were always vertically aligned and reversed their direction at the same moment. The students observed the movie freely. Their observing distance differed from about 5 to 12 m. They chose one of the three alternatives for each question. As for the spatial question, they judge the position of the target bounce at the right end relative to the references, either "undershoot" (the target bounced at the leftward position relative to the references), "overshoot" (the target bounced at the rightward position), or "aligned". As for the temporal position, they judged the relative timing of the target bounce, either "earlier" (than the references), "later", or "simultaneous."

The result is shown in Table 1. Most observers perceived that the target bounced at the "overshoot" position (16 of 20) and "later" timing (15 of 20) than the references. Taking in to account the various observing distances, this perceptual phenomena should have considerable robustness. Given this result, we next measure the more detail spatial and temporal characteristics of this phenomena.

Table 1. Numbers of Judgments of preliminary observation ($n = 20$).

Question	Judgment of the target (relative to the reference)		
Spatial	Undershoot	Aligned	Overshoot
	1	3	16
Temporal	Earlier	Simultaneous	Later
	2	3	15

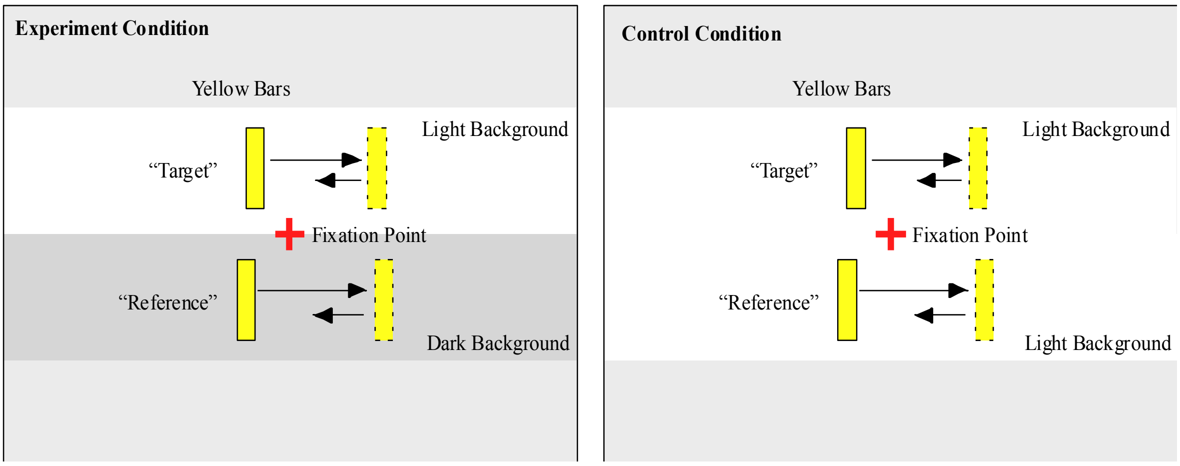


Fig. 2. Display configuration for the main experiment. There are two conditions, experimental and control. The control condition contains two identical bars on the same light background of the target bar of the experimental condition. The spatiotemporal relationship between the target and the reference was varied under the method of constant stimuli. The start, bounce, and end points of each stimulus randomly fluctuated within small range in order to reduce the predictability from stimulus regularity.

Experiment

Method

In the experiment, we changed the stimulus configuration as shown in Figure 2. The shape of moving stimuli was changed to vertical bar in order to measure accurate displacement. The fixation point was added at the center of the screen. One target and one reference moved above and below the fixation respectively. In each trial, both bars appeared and moved rightward for about 1,000 msec and then reversed their direction ant moved about 500 msec before vanished. The speed of movement was fixed at 2.4 deg/sec. The target and reference bars were yellow and their luminance was 60.7 cd/m^2. The luminance of the backgrounds was either 56.5 cd/m^2 (light) or 6.3 cd/m^2 (dark).

Four observers participated in the experiment, including one author. They were requested to judge the spatial or temporal relationship between upper and lower bars (naïve observers were not notified which was the target). We asked the observers to make a 2AFC judgment. The spatial question was "Did the upper bar bounced at whether overshoot or undershoot position relative to the lower?" and the temporal question was

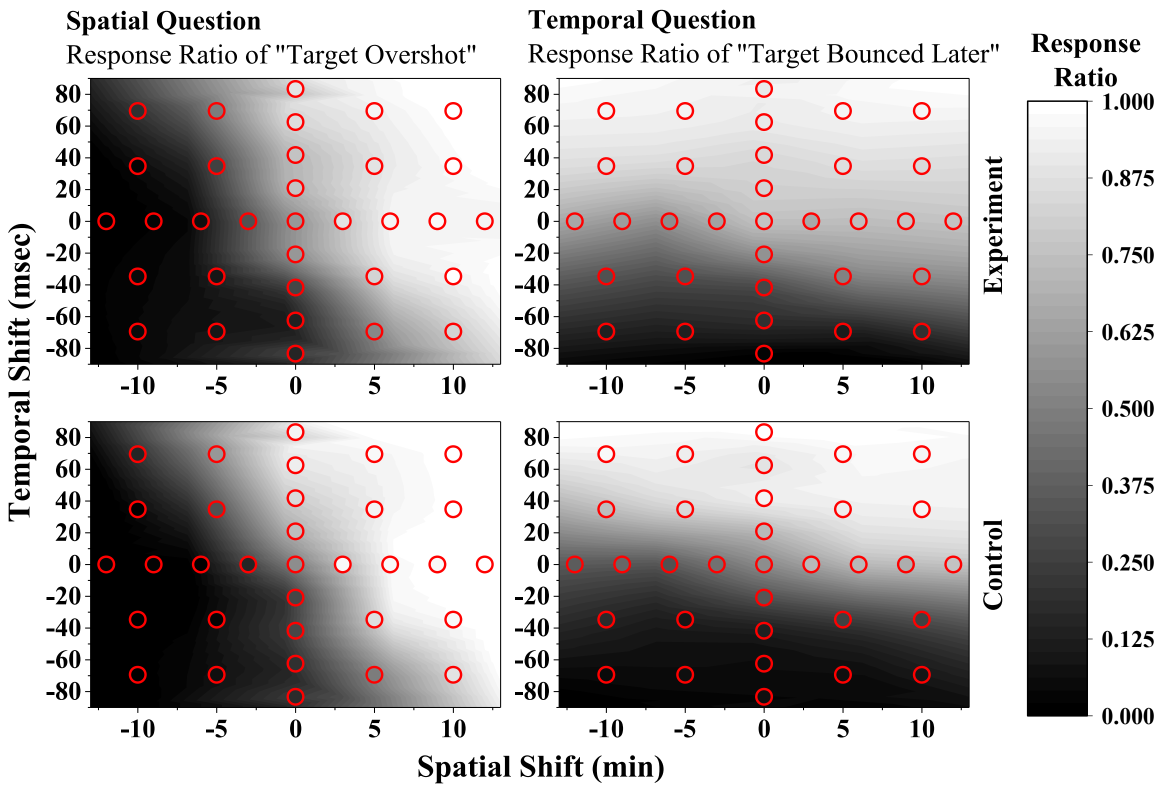


Fig. 3. Response ratio of each question and condition as grayscale maps. The gray level scale at the right side denotes the ratio of each judgment. Brighter area of each panel shows where the correspondent judgment ratio was high. Circles in each panel indicate the actual measured values. Negative values of spatial and temporal shift mean that the target bounced at the left side of the reference and earlier than the reference, respectively.

"Which bar bounced later, upper or lower?" We used the method of constant stimuli, varying both spatial and temporal shift of the target relative to the reference. So the target bounced at various relative position and timing conditions, including aligned and simultaneous one. There was a total of 33 combinations of spatial and temporal shifts and the observers repeated 20 judgments in quasi-random manner for each spatiotemporal condition.

In addition to the experimental condition, we introduced the control condition where the reference was on the same background as the target. In this condition the two bars were completely identical.

All of the above conditions and repetitions were conducted for two questions, spatial and temporal relationships. Trials with these two questions were separated into different blocks.

Results and Discussion

Overall Spatiotemporal Profiles: Differences in Question Types

All data from four observers were merged to one large data pool. For the spatial question, the ratio of "the target overshot" judgment was calculated for each spatiotemporal shift condition. Similarly the ratio of "the target bounced later" was obtained. Figure 3 shows the results of four stimulus/question combinations. For each panel, the horizontal and vertical axes are the spatial and temporal shift of the target respectively. The black-to-white level indicates the ratio of the judgment.

Comparing the results of the question types, there are differences in the distribution patterns of white areas. The left panels in Figure 3 are the result of spatial question. There is a tendency of diagonal pattern of the black and white boundary. This suggests that spatial judgment is dependent on the temporal shift. Regardless of the experiment/control conditions, the bar that bounced later tend to appear overshoot the bounce position relative to the earlier bounced bar.

In contrast, the black and white boundaries of the temporal question results (right panels) look almost horizontal, suggesting that the temporal judgment seems relatively independent of the spatial shift.

This difference of dependency let us think that the temporal relationship of the visual events is detected and processed almost in the manner of "as-is," but the spatial relationship might be processed taking the temporal relationship into account. It sounds as if the visual system interprets that the reason that one object bounced later than the other is that it went farther until the reversal occurred.

Spatial and Temporal Shift of the Experimental Condition

In order to see the spatial "overshoot" and temporal "delay" of the target clearly, we replotted the data of relevant points into the x-y plots (Figure 4). We extracted nine points on the x-axis from two spatial question results. These sets are the experimental and control conditions that only spatial shift was varied, keeping temporal shift zero (simultaneous bounce). The nine points on the y-axis from right panels make two sets of temporal shift results from experimental and control conditions in a same manner.

In both graphs, the data of experimental condition shifts leftward relative to the control condition. This means that the low-contrast bar was perceived to bounce at "overshoot" position and later timing relative to the high-contrast bar. Although the amounts of shift are small (rough estimations of the 50% point are 1 min arc for spatial overshoot and 20 msec for temporal delay), individual results were basically consistent with this overall results.

These results suggest that the spatial and temporal aspects of a moving stimulus can be distorted at the same time. In addition, according to the response pattern in Figure 3, the temporal aspects might be processed in a relatively independent manner of the spatial processing, but spatial aspects might be dependent on the temporal processing. The question is, however, the estimated spatial misplacement is much smaller than the amount that can be converted from the estimated time delay and the stimulus speed. The processing of spatial (mis)localization might be other than a simple conversion from time to space.

As can be seen in Table 1 and Figure 4, the bouncing position of low-contrast

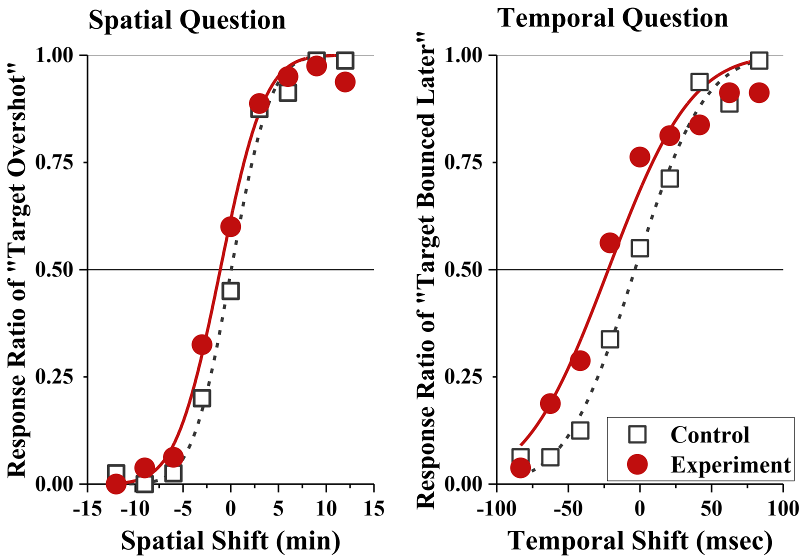


Fig. 4. Spatial and temporal misalignment. Data are extracted from Figure 3. For the left panel (spatial question), the points on the y axis (temporal shift = 0) from the left two panels of Figure 3 are plotted, the response ratio is the vertical axis here. For the right panel (temporal question), the points on the x axis (spatial shift = 0) from the right two panels of Figure 3 are plotted, with the y axis of Figure 3 is transferred into the x axis here. The sigmoid curves are the cumulative normal functions fitted to each data set.

stimulus consistently appear to "overshoot." This result seems contradictory to a simple concept that the slow-speed perception of low-contrast stimulus leads to its backward mislocalization backward at each moment. There is no reason for a forward mislocalization to happen only at the moment of the bounce.

Meanwhile, the mislocalization of flash-lag effect can be enlarged by decreasing contrast (Kanai, Sheth, & Shimojo, 2004). Though the spatial characteristics of our results look consistent with theirs, our results have significance in showing that the temporal delay co-occurs with the spatial mislocalization. Further investigation will be necessary to examine whether the temporal aspect of our results can also be explained by their theory.

References

Anstis, S. (2001). Footsteps and inchworms: Illusions show that contrast affects apparent speed. *Perception*, 30(7), 785–794. `http://doi.org/10.1068/p3211`

Kanai, R., Sheth, B. R., & Shimojo, S. (2004). Stopping the motion and sleuthing the flash-lag effect: spatial uncertainty is the key to perceptual mislocalization. *Vision Research*, 44(22), 2605–19. `http://doi.org/10.1016/j.visres.2003.10.028`

Nijhawan, R. (1994). Motion extrapolation in catching. *Nature*, 370(6487), 256–7. `http://doi.org/10.1038/370256b0`

Thompson, P. (1982). Perceived rate of movement depends on contrast. *Vision Research*, 22(3), 377–380. `http://doi.org/10.1016/0042-6989(82)90153-5`

MULTI-STABLE MOTION PERCEPTION IN THE POLKA DANCE STIMULUS

Hiroaki Yano[1], Yoshitaka Nakajima[2], Kazuo Ueda[2], Tatsuya Yoshizawa[3], and Gerard B. Remijn[2]
[1] *Graduate School of Design, Kyushu University, Fukuoka, Japan*
[2] *Department of Human Science/Research Center for Applied Perceptual Science Kyushu University, Fukuoka, Japan*
[3] *Research Laboratory of Affective Design Engineering Kanazawa Institute of Technology, Japan*
<HY: zyagi8e.u@gmail.com; GBR: remijn@design.kyushu-u.ac.jp>

The "Polka dance" stimulus is a multi-stable stimulus consisting of two identical bars that move in circular motion, with acceleration towards and deceleration away from the bars' point of coincidence (Fig. 1A). With steady fixation, the bars can be seen as moving through each other (streaming), as bouncing off of each other (bouncing), or as moving in circular motion pivoting around an imaginary midpoint (clockwise or counterclockwise rotation). Here we show that these four percepts appear in fairly similar proportions when the bars have an identical size and move with the same velocity (Fig. 1B). When bar sizes are unequal, the smaller bar appears in the background of the larger bar and rotation of both bars in unison is not perceived. Instead of streaming, bouncing or rotation, viewers often perceive different percepts in this case, such as figure-eight movement (Fig. 1C). As a more complex version of the streaming-bouncing stimulus, the Polka dance stimulus provides a way to study visual grouping in the perceptual translation of 2D motion images into 3D percepts.

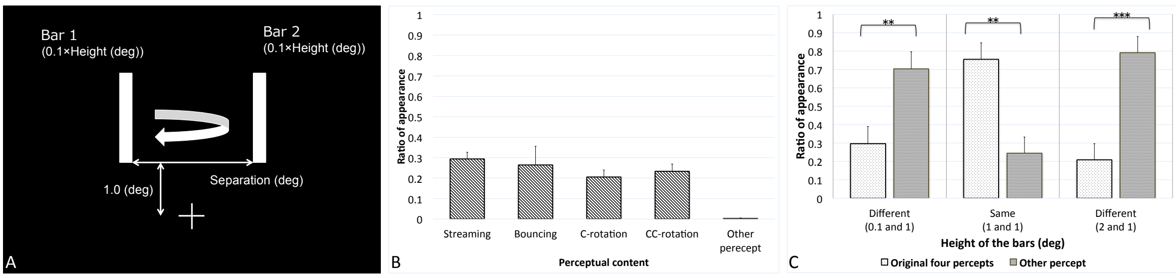


Fig. 1. (A) The Polka dance stimulus. (B) Ratio of appearance of streaming, bouncing, clockwise rotation and counter-clockwise rotation observed in the Polka dance stimulus when the bars have the same size. Error bars indicate the standard error of the mean. (C) The light-gray bars show the ratio of the cumulative appearance of streaming, bouncing, clockwise rotation or counter-clockwise rotation ("original four percepts") when the bars in the Polka dance stimulus have the same height (1 and 1 deg) or different in height (0.1 and 1 deg, or 2 and 1 deg). The dark bars show the ratio of appearance of percepts other than the original four percepts under these stimulus conditions. Error bars indicate the standard error of the mean. (**: p <0.01, ***: p <0.005)

PRE-PRESENTATION OF RANDOM MOTION REDUCES THE LATENCY OF VECTION

Jing Ni[1], Hiroyuki Ito[2,3], Masaki Ogawa[2], and Shoji Sunaga[2,3]
[1] *Graduate School of Design, Kyushu University, Fukuoka, Japan*
[2] *Faculty of Design, Kyushu University, Fukuoka, Japan*
[3] *Research Center for Applied Perceptual Science, Kyushu University, Fukuoka, Japan*

In this study, we aimed to understand what the latency of vection onset reflects. Before presenting a vection inducing stimulus, one of three kinds of visual stimuli was presented, i.e. a fixation cross (control), static random dots, or dots moving in random directions. The duration was 1 s, 2 s, 4 s or 8 s. After that, upward or downward motion of random dots that could induce vection was shown to participants. The latency of vection onset was measured. The results showed that the latency was shortest when the random motion was presented before presenting the vection inducing stimulus. There was no significant difference in latency between the control and the static random-dot conditions. Because it is not plausible that pre-presentation of random motion by itself reduces the mismatch between visual and vestibular signals, the results may indicate that pre-activation of the visual system contributes to visual dominance at the subsequent presentation of a vection inducing stimulus. (Supported by the JSPS)

INHIBITION OF VECTION BY OBJECT GRASPING MOVEMENT

Masaki Mori[1,2] and Takeharu Seno[3]
[1] *Graduate School of Media and Governance, Keio University, 5322 Endo Fujisawa 252-0882, Japan*
[2] *Japan Society for the Promotion of Science, 5-3-1 Kojimachi, Chiyoda-ku Tokyo 102-0083, Japan*
[3] *Department of Human Science/Research Center for Applied Perceptual Science Kyushu University, Fukuoka 815-8540, Japan*

It has been reported that visually induced self-motion perception (known as vection) is facilitated by holding a glass full of water (Seno et al., 2016). The present study investigated whether vection could be modified by object grasping movement. Twenty-five university students were asked to do one of the following four types of hand movements while they were viewing a radial optic flow; i.e. grasping a hand-gripper strongly, just grabbing a hand-gripper, clenching fist strongly and without having anything in their left hands. Participant's task was to keep pressing a button with their right hands while they were perceiving vection. After the stimulus presentation, they estimated subjective vection strength by 101-rating scale (0–100). As a result, it was revealed that vection was inhibited by grasping a hand-gripper strongly. We named this manipulation for inhibiting vection "the Gripping method".

DISSOCIATION BETWEEN PERCEPTION AND ACTION IN INDUCED MOTION: A COMPARISON BETWEEN TENNIS PLAYERS AND NOVICES

Yasuhiro Seya[1] and Hiroyuki Shinoda[2]
[1]*Faculty of Human Informatics, Aichi Shukutoku University*
2-9 Katahira, Nagakute, Aichi 480-1197, Japan
[2]*College of Information Science and Engineering, Ritsumeikan University*
1-1-1, Noji-higashi, Kusatsu, Shiga 525-8577, Japan

Abstract

We investigated the dissociation between perception and action of tennis players and novices by measuring the induced motion (IM), that is, the illusory motion of a fixated object induced by background motion. In two experiments, participants viewed a horizontally moving target within a vertically moving random-dot pattern. In Experiment 1, participants reported the perceived direction of the target's motion during the stimulus presentation using a joystick. In Experiment 2, participants pointed to the last position of the target immediately after the target and background disappeared. In Experiment 1, we found a larger magnitude of IM in the tennis players than in the novices. The magnitude of IM became larger with increasing velocity of background motion. Experiment 2 showed accurate performance in pointing to the target's last position. No differences were found between the groups or between the background motion velocities. These results provide evidence in support of the dissociation between perception and action.

In daily life situations, we must perceive visual information quickly and accurately for choosing and executing appropriate actions. Contrary to this fact, many studies investigating the effects of visual illusion on perception and action have reported that different visual information contributes to perception and action (e.g., Bridgeman, Kirch, & Sperling, 1981). For example, Bridgeman et al. (1981) investigated this topic by using induced motion (IM), i.e., the illusory motion of a fixated object induced by background motion. In their study, participants manually pointed to a stationary target which appeared moving due to IM. Their results indicated that pointing performance was immune to IM.

In the present study, we investigated the dissociation between perception and action by focusing on individual differences in IM. One of the mechanisms generating IM is fixational effort to suppress the reflexive eye movements in response to background motion, i.e., optokinetic nystagmus (OKN) (e.g., Post & Leibowitz, 1985; Seya, Ishihara, & Imanaka, 2015). According to this explanation, the magnitude of IM reflects the degree of fixational effort suppressing OKN (in response to moving background) to remain fixated on the target. In the context of this explanation, Seya and Mori (2007) assumed that karate athlete would perceive larger magnitude of IM because their regular karate training may lead them to exert a larger degree of fixational effort than novices. Their results supported this prediction.

In this study, we chose tennis players as our participants. Research has reported that tennis players demonstrate a distinctive pattern of fixation (Cauraugh, Singer, & Chen, 1993; Murray & Hunfalvay, 2017), and it can be expected that regular tennis training may have led such players to exert a larger degree of fixational effort, resulting in larger magnitude of IM in the players than in the novices. However, regarding the dissociation between perception and action, it can be expected that the pointing toward a target that is perceived as moving owing to IM is accurate irrespective of IM magnitude perceived by the participants.

Experiment 1

Method

Twenty individuals, 10 active tennis players (9 males and 1 female) and 10 under graduate students (10 males) participated in the experiment. The players had played tennis for a mean of 9.0 ($\pm$3.4) years. The novices had no experience with regular tennis training or any other sports over the past 3 years. The participants gave written informed consent to this experiment.

A personal computer (Apple Mac Pro Early 2009) was used to control the experiment and generate stimuli that were presented onto an LCD display (Green House GH-LCT22B) with a refresh rate of 60 Hz. The size of the display was 47.7 cm $\times$ 26.8 cm in width and height. Stimuli were viewed binocularly from a distance of 50 cm. The stimuli used were similar to those used in previous studies (Seya, et al., 2015). In the display, a target and an inducing stimulus were presented on a black background (luminance 0.23 cd/m^2). The target was a cross subtending 1° $\times$ 1° and presented in red (luminance 32.7 cd/m^2) on the display subtending 51° $\times$ 30° The target moved either rightward or leftward in the middle of the display at a velocity of 6°/s. The inducing stimulus was a random-dot pattern consisting of 300 white dots (luminance 182 cd/m^2). A single dot was 1.5° in size. The inducing stimulus moved either upward or downward at a velocity of either 6°/s or 24°/s. Participants' responses were recorded by a joystick (Mad Catz, MC-PS37) at a sampling rate of 60 Hz.

The experiment was conducted in a dark room. Participants sat on a chair with their head fixed on a chinrest while viewing the display. After 3 min of adaptation to the dark room, participants practiced the task for several trials until they were familiar with it, after which the experimental session followed. At the beginning of each trial, a stationary target was presented at the center of the display. After the participants pressed a key, the target jumped 15° to the left or right from the center of the display and started to move horizontally. An inducing stimulus was also presented and started to move vertically. When the target reached 15° to the opposite side from the center of the display, it jumped to the initial location and moved horizontally again. Thirty seconds after the onset of the target and inducing stimulus motions, the target disappeared and only a stationary inducing stimulus was presented for 5 s. The participants' task was to adjust the slant of the joystick to match the perceived direction of the target's motion path while accurately pursuing the target. On all the trials, the angular deviation of the joystick from the horizontal was recorded as a measure of IM (Seya et al., 2015).

There were two blocks of 20 trials: 5 trials for each inducing stimulus velocity (6 and 24°/s), direction (upward and downward), and target motion direction (rightward and leftward). In one block, the target always moved rightward, while it moved leftward in the other block. The order of the two blocks was randomized across participants. The velocity and direction of the inducing stimulus were randomly selected for each trial.

For the joystick data, we first divided the data into three 10-s bins and calculated mean angular deviation after removing data that were more than 2 standard deviation away from the mean.

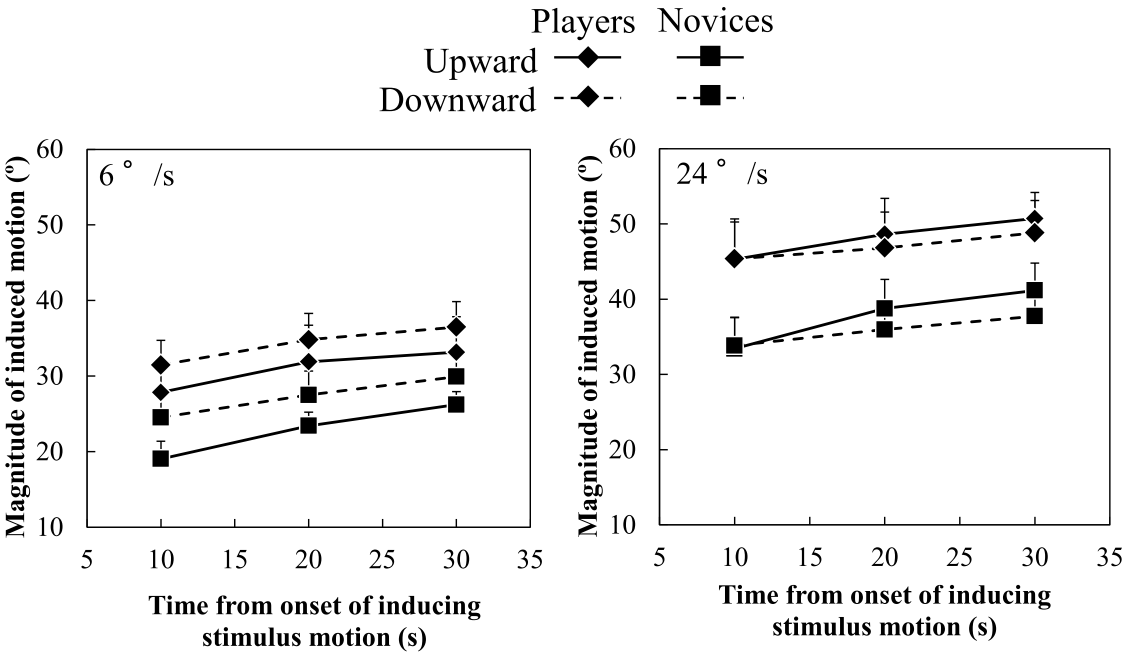


Fig. 1. Mean magnitude of induced motion (IM). Error bars indicate standard errors from the mean.

Results and Discussion

Figure 1 shows the mean magnitude of IM. As shown in the figure, IM magnitude was larger in the tennis players than in the novices. IM magnitude was larger during the latter part of the stimulus presentation than during the earlier parts. 2 (group) × 2 (velocity) × 2 (direction) × 3 (interval) ANOVA showed significant main effects of group, $F(1,18) = 4.42, p < .05$, velocity, $F(1,18) = 57.09, p < .01$, and interval, $F(2,36) = 31.08, p < .01$. There was an interaction of velocity × direction, $F(1,18) = 13.47, p < .01$. Multiple comparisons for the effect of interval showed significant differences between any pairs of intervals (all $ps < .05$). Subsequent analyses of interaction showed a significant simple main effect of velocity in both directions [upward, $F(1,36) = 70.34, p < .01$; downward, $F(1,36) = 30.81, p < .01$], while no significant effect of direction at both velocities was found.

The results showed larger magnitude of IM in the tennis players than in the novices. This finding is in agreement with the view that sports athletes, whose fixation stability is considered to be enhanced through training, perceive larger IM than novices do (Seya & Mori, 2007). Regarding the effects of stimulus velocity, we found larger IM with increasing stimulus velocity, which is consistent with previous studies (Lott & Post, 1993; Seya et al., 2015, but see Seya & Mori, 2007). Furthermore, we also found larger IM during the latter part of stimulus presentation than during earlier parts of stimulus presentation, which is consistent with previous findings (Lott & Post, 1993; Seya et al., 2015).

Experiment 2

Method

The method was identical to that of Experiment 1, with the following exceptions. First, the target was presented for 26 to 30 s, after which the display was replaced with a blank. The presentation duration was randomly selected for each trial. Second, the task of participants was to pursue the target accurately, and to point the location of target immediately after the

target (and random dots) disappeared, using their index finger of right hand. After the pointing was made, stationary dots were presented for 5 s. Pointing performance was measured in all the trials. There were 20 participants: 10 athletes and 10 novices. All the participants participated in Experiment 1.

Results and Discussion

For the pointing data, we calculated the position errors based on the position of the target offset and averaged them. Because we expected the horizontal position error to depend on the pursuit direction, we analyzed the data separately for each pursuit direction. Figure 2 shows mean position errors for each pursuit direction. As shown in the figure, pointing was biased rightward during rightward pursuit while it was biased leftward during leftward pursuit. Regarding the vertical position errors, pointing was biased toward the IM direction, i.e., direction opposite to inducing stimulus motion. Note that the position errors were very slight in both axes.

We converted horizontal errors into positive and negative values when they were

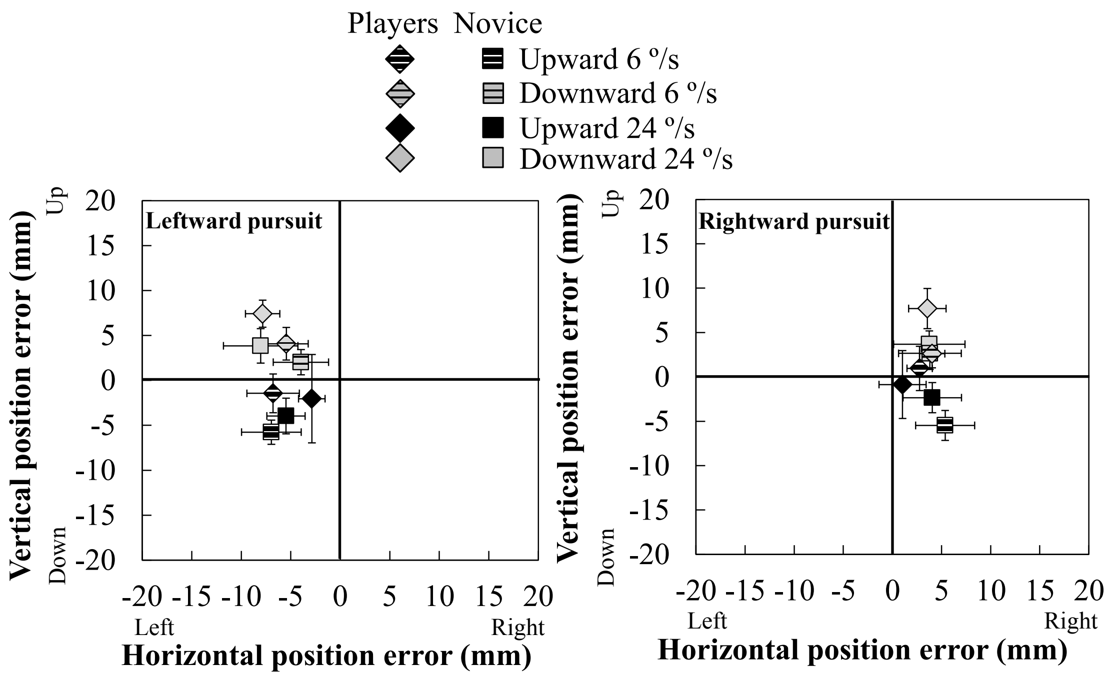


Fig. 2. Mean position errors of pointing performance in Experiment 2. Error bars indicate standard errors from the mean.

in and against the pursuit directions, respectively. A 2 (group) $\times$ 2 (pursuit direction) $\times$ 2 (velocity) $\times$ 2 (direction) ANOVA showed only an interaction of velocity $\times$ direction, $F(1, 18) = 8.20, p < .05$. A subsequent analysis showed a significant simple main effect of direction at 24°/s, $F(1, 36) = 5.78, p < .05$. For vertical position errors, we converted them into positive and negative values when the errors were in and against the IM direction, respectively. We then conducted a $2 \times 2 \times 2 \times 2$ ANOVA, which showed only a significant interaction of velocity $\times$ direction, $F(1, 18) = 12.61, p < .01$. Subsequent analysis showed a significant simple main effect of velocity in the downward stimulation, $F(1, 36) = 7.07, p < .05$.

The results showed no systematic differences by the groups or the stimulus velocity, which is not consistent with the results of Experiment 1. It should be noted that in this experiment, the target disappeared in the latter part of the stimulus presentation, during which a larger magnitude of IM than during the earlier parts was found in Experiment 1. These results strongly support the idea of the dissociation between perception and action.

General Discussion

The present study examined the dissociation between perception and action by focusing on the individual differences in IM. Experiment 1 showed larger magnitude of IM in the tennis players than in the novices. On the other hand, Experiment 2 showed no effects of group or stimulus velocity on pointing performance. Similar dissociation between perception and action was found for the effect of the stimulus velocity. Taken together, the present results clearly support the notion that different visual information contributes to perception and action (e.g., Bridgeman, 1981).

Acknowledgements

This research was supported by JSPS KAKENHI Grant Numbers JP17H00875, JP17K00286 and JP15K00216.

References

Bridgeman, B., Kirch, M., & Sperling, A. (1981). Segregation of cognitive and motor aspects of visual function using induced motion. *Perception and Psychophysics*, 29, 336–342.

Cauraugh, J., Singer, R.N., & Chen, D. (1993). Visual scanning and anticipation of expert and beginner tennis players. In S Serpa (Ed.), *Proceedings : VIII World Congress of Sport Psychology* (pp.336–340). Lisbon: International Society of Sport Psychology.

Lott, L. A., & Post, R. B. (1993). Up-down asymmetry in vertical induced motion. *Perception*, 22, 527–535.

Murray, N. P. & Hunfalvay, M. (2017). A comparison of visual search strategies of elite and non-elite tennis players through cluster analysis. *Journal of Sports Sciences*, 35, 241–246.

Post, R. B., & Leibowitz, H. W. (1985). A revised analysis of the role of efference in motion perception. *Perception*, 14, 631–643.

Seya, Y., Ishihara, M., & Imanaka, K. (2015). Up-down asymmetry in vertical induced motion and optokinetic nystagmus. *Attention, Perception, & Psychophysics*, 77, 220–233.

DISCRIMINATION OF VEHICLE SPEEDS BY STANDING WATCHERS

Bruno de Araújo Faria and Ricardo Kamizaki
Federal University of Juiz de Fora/MG–Brazil
<bafaria@terra.com.br, rkz57@hotmail.com>

Abstract

The perception of speed and discrimination of this have been little studied during the development of psychophysics, therefore few jobs are found in the literature, less common still those who use roads and real vehicles. In this study, 50 volunteers of both sexes were asked to determine the velocity of a real vehicle moving in four different speeds, namely 20, 30, 40 and 50 km/h, while the volunteers remained stationary. The mean results and Pearson's correlation coefficients (0.99) demonstrate that there's great accuracy in determining the vehicle speeds with no apparent influence of the speed applied, participant sex, nor with the fact that this have or not driver license. These results showed that speed is also perceived as a primary characteristic of the human perceptual system, which was also corroborated by obtaining the empirical Stevens's Power Law (exponent = 1.26) for the variable studied.

The most complex human perceptual system is the vision. The main function of human vision is to construct a representational image of environmental objects from information extrinsic to the organism that are transformed into nerve impulses. This information of the environment, which are in the case of the vision, essentially: contrast, movement, details, shape, color, depth, etc. are compiled, codified and analyzed in cortical processes, being later associated to the memory and also to the information received by the other senses and then being synthesized resulting in the world we perceive (Araújo, 2014).

The perception of movement is one of the most important and fundamental characteristics of the human visual system allowing the effective interaction of individuals with objects and people and also their own displacement in space. This characteristic, acquired throughout the human evolutionary process, guaranteed the survival of the species since the perception of predators and other environmental threats was fundamental.

In spite of the considerable capacity of the human perceptual system to detect movement patterns, and this fact corroborated by the large number of studies related to the subject in the literature, speed perception and discrimination has not been the subject of many human perceptions studies (Costa, 2011). However, some studies suggest that speed perception is a primary feature of the human perceptual system, as well as time and space, and not just a derivation of these (Lappin, Bell, Harm & Kottas, 1975). One could think of this a priori by the effect of the physical concept of velocity itself as being the relation between these two variables (velocity = distance/time). Although we take these three dimensions of human perception as primary, we can't rule out the existence of the relationships between these variables, for example, in simple everyday tasks such as crossing a street, as demonstrated by Hoffmann (1994). Corroborating the idea that speed perception is a primary feature of the human perceptual system, Recarte & Nunes (1996) made such an assertion based on the conclusions of their studies of estimation and speed production by drivers because these researchers didn't find differences attributable to gender and driving experience.

With the technological development and the arrival of machines and automotive vehicles, not only the perception of the movement materialized as something fundamental, but also the importance of the discrimination of the objects speed movement was increased. After all, as Raghuram, Lakshminarayan & Khanna (2005) argue, dealing with high speeds and estimating them is essential in vehicular steering tasks. This importance can be more easily understood when dealing specifically with pedestrians and drivers, in other words, the former need to have an intuitive notion of the speed of movement of vehicles to cross, for example. On the other hand, drivers among the various visual tasks required to drive a vehicle should also be attentive to the speed of movement of pedestrians, cyclists and especially in estimating the speed of other vehicles which, as pointed out by Scialfa, Guzy, Leibowitz, Garvey & Tyrrel (1991), would be the most important task in driving, all this in order to avoid traffic accidents, many of them with fatal consequences.

According to what has been said above, although the perception of movement itself has, over the years, attracted the interest of researchers from different areas, specifically the discrimination of the objects speed movement has been neglected and few studies are found in the Literature, such as the research by Scialfa et al (1991), Raghuram, Lakshminarayan & Khanna (2005), Schutz, Billino, Bodrogi, Polin, Khanh, & Gegenfurtner (2015) and Milosevic & Milic (1990) who set out to investigate the relationships between speed discrimination and age and sex of participants. In these studies, the participants were inside the vehicles and obtaining the data consisted in estimating the speed of the vehicles in which they traveled or others that traveled.

All the previously related factors demonstrate the importance of the experimental arrangement adopted in this study, that is, in the natural environment and with the use of a real motor vehicle and not just a simulation. Another important aspect of this study is the fact that, in this case, we have stopped observers in relation to the mobile and not inside the vehicles as in these cited studies, giving the participants the position of inertial vehicle movement testimonies and not movement producers or localized testimonies inside the cabinet.

Methodology

Participants: Fifty volunteers were recruited, 25 males and 25 females, ranging in age from 19 to 51 years (mean = 24.55 years), 18 of whom had a driver licence and the rest did not. All participants had good visual acuity and those who stated that they needed corrective lenses were using it at the time of the experiment.

Methodology: The experiment was carried out on an open area, always in daytime, with good weather and dry track in one way of the Federal University of Juiz de Fora. This used stretch of the road was rectilinear, with good visibility in both directions of movement, adequate horizontal signaling and counted on a track divided in two bands of opposite directions. In order to collect data, the participants remained in small groups for each battery of tests performed on the sidewalk throughout the experiment and were instructed to note the presumed speed of the vehicle when it passed a pre-established landmark and visible to all the participants. The vehicle used was a touring car, Citröen, model C3 Attraction, year of manufacture 2015, model year 2016, in red color and was in mechanical conditions suitable for circulation. The speed printed on the vehicle was visually controlled by the driver throughout the experiment.

Four velocities were chosen for this study, being these 20, 30, 40 and 50 km/h. The driver printed the necessary acceleration in the stretches before the reference point of speed note presented to the participants and when the speedometer reached the value to be studied, the acceleration was kept constant until the passage through said point, in this way, described procedure was repeated 4 times for each group of participants.

Results and discussion

The analysis of all the data obtained in the different batteries of the experiment was done using adequate software (Excel®) to calculate the arithmetic means, standard deviations, and also the Pearson's correlation coefficients (r^2). The results for the whole sample set are shown in Table 1.

Table 1. Mean, standard deviation and relative error for the whole sample.

Statistical parameter	Speed (Km/h)			
	20	30	40	50
Mean (Km/h)	16.20	28.24	38.39	52.94
Standard deviation (Km/h)	5.77	5.32	9.55	10.33

Observing the data from Table 1 it is observed that there's a great proximity between the calculated mean values and each of the real values of the velocities employed in this study, especially if the standard deviations are taken into account. This is in line with the Law of Intensity of Stimulus or Law of Magnitude, which says that increasing the intensity of a stimulus tends to increase the response to it, and in this case specifically increases the accuracy of the response. The Pearson's coefficient r^2 calculation to verify the correlation between the actual velocities employed and the means obtained in the experiment, whose value was 1.00 (Figure 1), corroborates the above statement demonstrating the strong correlation between the given responses and the presented stimulus.

The results of the sex-separated sample reveal the same general trend discussed above for the whole sample, that is, means close to the actual values for both sexes, considering the calculated standard deviations evidencing again the Law of Magnitude, as shown in Table 2.

Also in this case, Pearson's r^2 was calculated for males, female and also in the comparative between both, whose values were respectively 0.99; 1.00 and 0.99. By observing the data set, a strong positive correlation was observed between the real values of the velocities used with the mean velocities estimated by each of the sexes considered separately and also between the means of both sexes when compared to each other

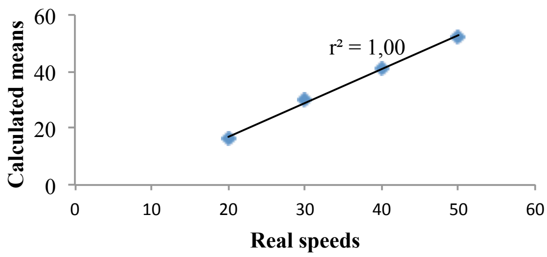


Fig. 1. Pearson's correlation coefficient for the calculated means of the whole sample.

Table 2. Table showing mean, standard deviation and relative error for the sample separated by sex (M = male and F = female).

Statistical parameter	Speed (Km/h)							
	20		30		40		50	
	M	F	M	F	M	F	M	F
Mean	16.12	16.27	27.96	30.00	35.76	40.92	52.12	52.50
Standart deviations	6.32	5.30	5.87	4.84	8.12	10.27	8.28	12.09

Table 3. Table showing mean, standard deviation and relative error for the sample separated by DL and non-DL.

Statistical parameter	Speed (Km/h)							
	20		30		40		50	
	DL	Non-DL	DL	Non-DL	DL	Non-DL	DL	Non-DL
Mean	16.33	15.71	28.06	28.39	36.61	39.32	50.11	54.61
Standart deviations	5.83	5.73	4.44	5.96	9.13	9.92	8.55	11.33

reaffirming the tendency exhibited by the sample when considered in its entirety, in the analysis previously made.

On this study, statistical analyzes were also performed considering another important variable: whether or not the individual has the driver's license (DL) and we will have two comparative groups again. The data are presented in Table 3.

Reiterating the conclusions obtained in the previous analyzes, one can again observe the values of the means with strong correspondence in relation to the real speeds used taking into account the standard deviations for each group. For the speed of 20 km/h (the lowest used in this study), The fact that the increasing magnitude of the stimulus causes an increase in response to it. Pearson's correlation coefficients were also calculated for these data sets by comparing the two groups with each other and also the means of the velocities estimated by the groups separately from the actual velocity values used. In all of these cases, the calculated Pearson's correlation coefficient ($r^2 = 1.00$) indicates that there are strong correlations, that is, having or not DL, the individuals were able to estimate the vehicle's movement speeds with a high degree of accuracy in comparison to the real value used in the experiment. Therefore, apparently, being a licensed driver isn't a variable that interferes with the ability to discriminate speeds.

Theory and Empirical Validation of the Stevens Power-Function Exponent

Still from a theoretical point of view, another interesting discussion for this study, which can offer a satisfactory explanation for the results presented here and at the same time corroborate the discussions here, is the calculation of the exponent of the power function of the speed perception. The power-function created by Stanley Smith Stevens is a mathematical function that relates the perceptual responses of the individuals to the magnitude of the real stimulus, that is, relates the subjective experience of the individual and the physical world (Da Silva & Macedo, 1982). The methods of measurement and calculation to obtain this relation start from the premise that it is possible to obtain direct measures of the sensory experience and, in this way, Stevens obtained the mathematical expression of the Power's Law:

$$R = KS^n \tag{1}$$

Where, R is the individual response to the presented stimulus; K is an arbitrary constant; S represents the real magnitude of the stimulus and n is a characteristic of the physical quantity studied. The most important term in equation 1 is the exponent—n—which is specific for each class of stimuli and its scalar value indicates the perceptual sensitivity of the individuals to a particular type of stimulus. Thus, an exponent less than unity ($n < 1.00$) indicates lower perceptual sensitivity for this stimulus and vice versa for exponent values greater than 1.00.

The table with values of n presented by Da Silva & Macedo (1982) doesn't present the value of the exponent for the stimulus studied in this work, that is, speed. However, it shows the values for visual distance (0.97 ± 0.22) and time (0.91 ± 0.18), so, by mathematical analogy one can easily deduce the theoretical exponent for the stimulus velocity considering the Laws of Classical Newtonian Mechanics, where the mean velocity (V) is the mathematical relationship between the variation of the position of a mobile in space, that is the distance traveled by this mobile (d) and the variation of the time elapsed for it (t). Thus, taking into account the law of propagation of errors for the calculation of the general standard deviation, we will have: $n(V) = 1.07 \pm 0.45$.

In order to obtain a theoretical discussion regarding the value of the power-function exponent obtained theoretically, it should be compared to the experimental value deduced from the data obtained in this work. However, to calculate the exponent value of this experiment, one must use the logarithmic form of equation 1, reducing it to a mathematically analogous form to a 1st degree equation, as suggested by Da Silva (1985). Thus, equation 1 becomes: $logR = nLogS + logK$ (equation 3). The equation of curve adjustment obtained by the logarithmic data of the values of the real velocities and velocities estimated from table 1 was $y = 1.26x - 0.42$ (equation 4). Thus, by analogy, the values of $K = 0.38$ and $n = 1.26$ are obtained.

Considering that the theoretical value calculated for the exponent of the power-function of velocity perception using the time and distance exponents available in the literature was 1.07 ± 0.45, as shown previously, it can be said that, taking into account the standard deviation, it is observed that this value is very similar to that obtained experimentally in this work, that is, 1.26. This fact corroborates the prediction of the Stevens's Law for the perception of the stimulus studied in this work, calculated indirectly using the time and distance data.

In addition to the above-described conclusion, one can additionally obtain the Stevens's Equation or Power Law in its complete mathematical form for the "velocity" stimulus by substituting the values of n and K in equation 1: $R = 0.38S^{1.26}$ (equation 5).With the equation in its integral form, we can obtain the graph of the resulting curve for the power function of the velocity, which shows the trend for all the possible values assumed by this stimulus along a continuum (Figure 2).

The mathematical analysis of equation 6 shows that at least in theoretical terms and according to the mathematical prediction of this psychophysical model there is a certain tendency of individuals to perceive speeds higher than those actually applied in relation to less intense stimuli in continuous series of presented stimuli. In other words, the greater the magnitude of the stimulus used, the greater the estimated value of the individuals for this stimulus, which also directly affects the values of the calculated

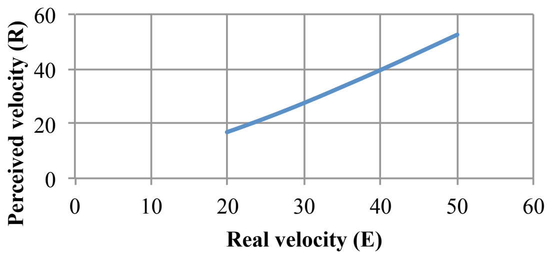


Fig. 2. Curve of the Stevens's equation for velocity—$R = 0.38S^{1.26}$.

experimental errors whose tendency is to increase as the intensity of the stimulus. The mathematical prediction described above can be easily observed by the simple analysis of the values available in Table 1. Thus, when the greatest stimulus was presented, that is, the velocity of 50 km/h, it was perceptibly perceived by the participants with a higher value—52.94 Km/h. This value is quite close to the theoretical prediction given by equation 6 which returns the value of 52.54 km/h. Analogously and inversely, the stimulus of lower intensity, that is, the velocity of 20 km/h was perceived as lower, 16.20 km/h, also close to the value provided by equation 6—16.56 km/h.

Conclusion

In this work, using a very simple experimental arrangement, allied to an original methodology, we sought to contribute to the increase of the knowledge of the psychophysics of human perception regarding speed. Thus, in view of the results of the velocity averages estimated by the participants obtained in the experiments, it was noticed a clear tendency of the individuals to estimate with a high level of accuracy the velocities of a mobile when the first ones are placed in the condition of inertial observers of the movement.

In addition to the conclusions based on the averages and errors previously discussed, the Pearson's correlations calculated presented quite significant results from the theoretical point of view, and in all cases of correlations studied, the coefficient reached values close to or equal to 1.00 (subsumed value on Statistics as a reference for a perfect correlation). Thus, these facts allow us to infer that, especially in the case of the two variables studied in this study, that is, the sex of the individual and his or her ability to drive, don't seem to interfere in the ability of these individuals to estimate the speeds of a vehicle in motion. Although other variables may be object of study in future works, preliminary data indicate that such capacity must be inherent in the human species itself that for thousands of years of evolution has adapted not only to the perception of movement itself, but also to discriminate the speed of displacement of objects in nature which is a very important and inherent characteristic of the movement itself.

All the above conclusions support the idea that speed discrimination is a primary characteristic of human perception, as well as time and space, as well as that observed in studies by Recarte & Nunes (1996), Raghuram, Lakshminarayan & Khanna (2005) and Schutz et al. (2015).

References

Araújo, A. R. M. (2014). *Fatores Sensoriais Visuais que Influenciam o Dimensionamento Subjetivo na Percepção de Tamanho: Um Estudo de Escalonamento Psicofísico.* Dissertação de Mestrado, Universidade de São Paulo, São Paulo, SP.

Costa, A. M. P. C. (2011). *Escalas psicométricas e propriedade escalar na percepção de velocidade em pombos.* Dissertação de Mestrado, Universidade do Minho, Braga, Portugal.

Da Silva, J. A. & Macedo, L. (1982). A função-poténcia na percepção—significado e procedimentos de cãlculo do expoente. *Arquivo Brasileiro de Psicologia,* 24(4), 27–45.

Hoffmann, E. R. (1994). Estimation of time to vehicle arrival—effects of age on use of available visual information. *Perception,* 23, 947–955.

Lappin, J. S., Bell, H. H., Harm, O. J. & Kottas, B. (1975). On the Relation Between Time and Space in the Visual Discrimination of Velocity. *Journal of Experimental Psychology: Human Perception and Performance,* 1(4), 383–394.

Milosevic, S. & Milic, J. (1990). Speed Perception in Road Curves. *Journal of Safety Research,* 21, 19–23.

Raghuram, A., Lakshminarayanan, V. & Khanna, R. (2005). Psychophysical estimation of speed discrimination. II. Aging effects. *Journal of the Optical Society of America,* 22(10), 2269–2280.

Recarte, M. A., & Nunes, L. M. (1996). Perception of speed in automobile: Estimation and production. *Journal of Experimental Psychology-Applied,* 2, 291–304.

Schutz, A. C., Billino, J., Bodrogi, P., Polin, D., Khanh, T. Q. & Gegenfurtner, K. R. (2015). Robust Underestimation of Speed During Driving: A Field Study. *Perception,* 44(12), 1356–1370.

Scialfa, C. T., Guzy, L. T., Leibowitz, H. W., Garvey, P. M. & Tyrrel, R. A. (1991). Age diferences in estimating vehicle velocity. *Psychology and Aging,* 6(1), 60–66.

A NEW AFFECTIVE MEASUREMENT WITH A JOYSTICK REVEALS TEMPORAL CHARACTERISTICS OF AUDIO-VISUAL INTERACTION IN FILM VIEWING

Masaharu Nomura and Hiroshi Arao
Taisho University, Japan

Since films consist of a series of audio-visual events that evolve along with time, affective responses toward them may also vary along with time. Current typical measurements, however, capture affective responses toward a movie stimulus as a whole at best. The purpose of the present study is to reveal temporal characteristics of audio-visual interaction in film viewing, utilizing a new affective measurement with a joystick. A pair of 70-second opening and action scenes was extracted from 2 films. Each of them was presented in audio-visual, audio-only, and visual-only modes. Twenty participants evaluated them by moving a joystick forward or backward as corresponding to the varying degree of interest (DoI). DoI values were measured at 60Hz and ranged from -400 (lowest interest) to 400 (highest interest). DoI values were generally higher in the audio-visual mode than those in the other modes across time segments and scenes. Furthermore, higher DoI values in the audio-visual modes could not be accounted for by even the sum of those in the audio-only and the visual-only modes, demonstrating temporal and quantitative characteristics of net audio-visual interaction effects. As such, our method revealed several unique patterns of audio-visual interaction. Fig. 1 and Fig. 2 show DoI values in the 3 presentation modes and the sum of those in the audio-only and the visual-only modes for Kingsman (2014) action scene and for Transporter 2 (2005) opening scene, respectively.

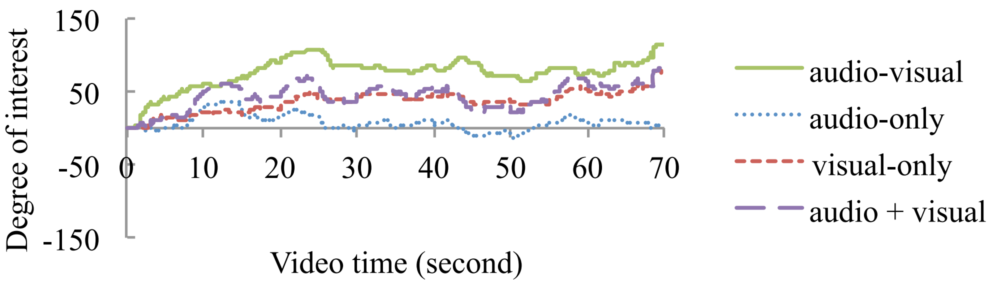


Fig. 1. DoI values for Kingsman (2014) action scene.

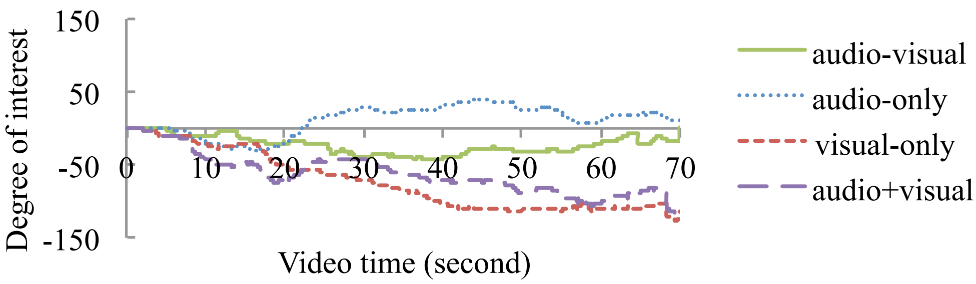


Fig. 2. DoI values for Transporter 2 (2005) opening scene.

THE EFFECTS OF SOUND FREQUENCY ON SOUND-INDUCED FLASH ILLUSION

Yu Masago, Takuma Matumura, Yuta Tamai, Takafumi Furuyama, Shizuko Hiryu, and Kohta I. Kobayasi
Neuroesology and bioengineering Laboratory, Doshisha University
Tataramiyakodani, Kyoutanabe-shi, Kyoto 610-0321, Japan
<kkobayas@mail.doshisha.ac.jp>

Abstract

When a flash stimuli is presented once and at the same time, a sound stimuli is presented twice, we perceive that the flash is presented as if it were presented twice. This phenomenon is called sound-induced flush illusion, and the mechanism of auditory-visual integration that causes this phenomenon has not been clarified yet. In order to understand the conditions required for the auditory-visual integration, we examined how sound frequency affects the illusion. In this study, we presented a pair of tone burst with singe flash, and frequency difference of the pair was systematically changed. As a result, when different frequencies were presented, the rate of occurrence of the illusion decreased. The result suggests that frequency processing could be involved in the auditory-visual integration.

We obtain multiple sensory information in everyday life and perceive the environment by integrating those information. An illusion sometimes occurs when contradicting visual and auditory information are integrated. One of such is called Sound-induced flash illusion, originally discovered by Shams and her colleague (Shams et al., 2000; Shams et al., 2005). In the phenomenon, single flash is perceived as flashing twice or more when multiple beeps are simultaneously presented. By examining what kind of stimulus condition is needed for Sound-induced flash illusion to occurs, the mechanism of audiovisual integration can be clarified. In this study, we studied whether the ratio of occurrence of Sound-induced flash illusion changes when auditory stimuli of different frequencies are presented with a visual stimuli at the same time.

Method

Experiment 1: Measuring Visual Temporal Resolution

Subjects were 5 males aged 23–26. All experiments were conducted in a soundproof room. We used a standard PC monitor for presenting a visual stimulus in the soundproof room. Subjects carried out experiments at a distance of 60 cm from the monitor. The visual stimulus is illustrated in detail in Fig.1. The fixation point was presented at the center of the monitor. Visual stimulation, a white circle, was shown at a distance of 21 mm below the fixation point. Visual stimulus was presented twice with duration of 8 ms, and 10 types of inter stimulation interval (ISI) between 8–80 ms (8, 16, 24, 32, 40, 48, 56, 64, 72 and 80) were used. Ten types of stimulus were presented 22 times in random order in one session, and each subjects conducted 10 sessions (= 220 trails). The subjects were asked to answer whether stimulus was seen once or twice. From the results, visual temporal resolutions (i.e., point of subjective equality: PSE) for each subject were estimated, and

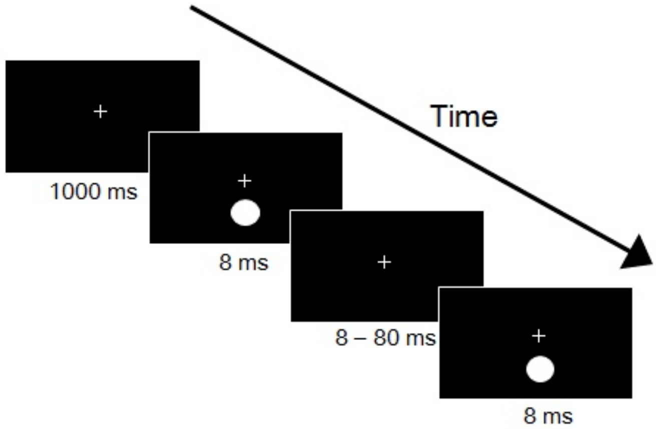


Fig. 1. Time course of visual stimulus.

the values was used in the later experiment.

Experiment 2: The Effect of Sound Stimulus Change on Audiovisual Integration

Same subjects as Experiment 1 were participated. Experimental 2 setting were also same as Experiment 1 unless stated otherwise. Sound stimulus was presented by a head phone placed next to the screen. The beginning of the visual and the auditory stimulus was synchronized. Visual stimulus was two types: single or paired. For pared visual stimulus, the inter stimulus interval was set to PSE of each subject determined in Experiment 1. The tone burst of 80 dB was used for auditory stimulation. The duration was 10 ms with the rise and fall time of 1 ms. The auditory stimulus was a constant frequency tone burst, always presented twice, and the interval between stimuli was 48 ms. The frequency of the stimulus ranged between 2000 to 4000 Hz divided by 1/6 octave: 2000, 2245, 2520, 2828, 3175, 3564, and 4000 Hz. We presented 13 combinations of tone bursts, at least first or second tone burst is 2000 Hz (see Table 1 for details). The 14 types of sound stimulus,

Table 1. Combinations of sound stimulus.

Sound stimulus	First sound [Hz]	Second sound [Hz]
Same frequency	2000	2000
Frequency raise	2000	2245
	2000	2520
	2000	2828
	2000	3175
	2000	3564
	2000	4000
Frequency fall	2245	2000
	2520	2000
	2828	2000
	3175	2000
	3564	2000
	4000	2000

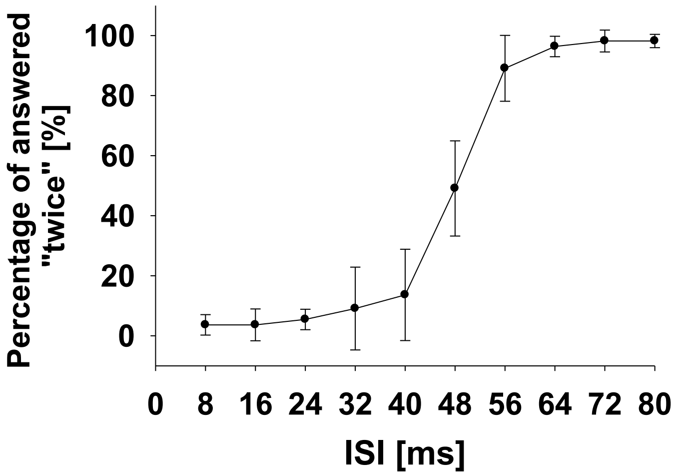


Fig. 2. Percentage of answered "twice" when presented pair of flashes with varying inter stimulus intervals(ISI) The abscissa is the ISI of the visual stimulus and the ordinate is the rate the subject answered as twice. Error bars are the standard deviation of 5 subjects.

the 13 combination and no-sound condition, were paired with two types of visual stimulus (single or paired flashes) to make 28 types of stimulus. Each stimulus types were presented in random order 15 times, resulting in 420 trials per subject. The subjects were asked to answer whether stimulus was seen once or twice.

Results and Discussion

Experiment 1: Measuring Visual Temporal Resolution

Figure 2 shows the average visual temporal resolution measured with varying visual stimulus interval. The PSE of each subject was defined as the ISI whose ratio of responding to visual stimulus twice is the closest to 50%. Individual difference was relatively small; PSE was estimated as 48 ms in 4 subjects and 40 ms in one subject.

Consideration on Audiovisual Integration

Figure 3 shows percentage of a single flash perceived as twice when it was presented with different combination of auditory stimulus. The result suggests that Sound-induced flash illusion in general occurred, because the percentage of subjects answering twice is high than zero in almost all stimulus conditions(Shams et al., 2005; Rosenthal et al., 2009). When the frequency of the first and the second sound is the same, the percentage at which the subject answered visual stimulus twice was the highest. The illusion occurs less when the frequency difference between the first and the second sound increases, the Sound-induced flash is not independent of sound frequency. In addition, when comparing Figuer 3a and Figuer 3b, the illusion occurrence is different between frequency-raise and frequency-fall condition. Frequency-fall stimulus evoked more illusion than corresponding frequency-raise stimulus. The result suggests that not only frequency diffence but also temporal frequency pattern is an important contributing factor for the illusion, and raise

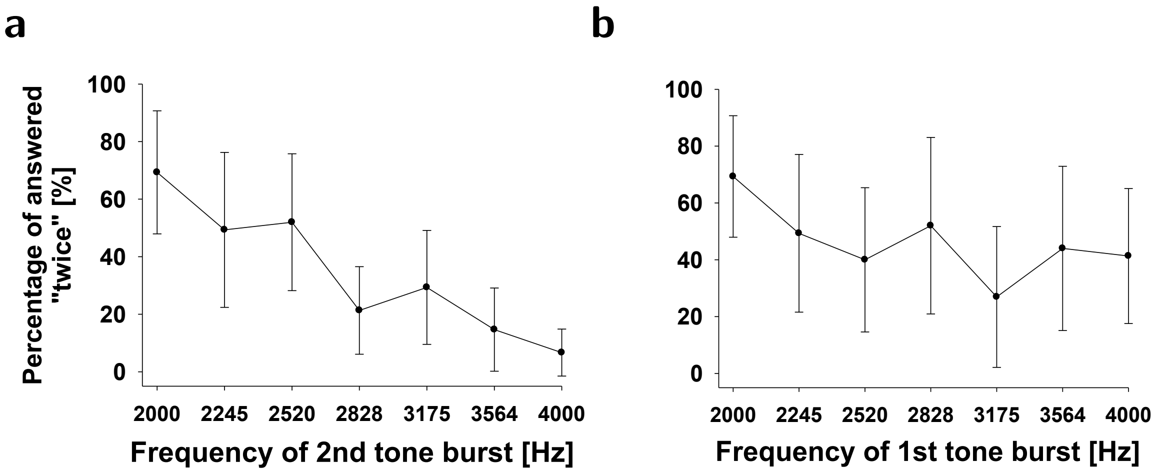


Fig. 3. Percentage of answering visual stimulus perceived as twice when single visual stimulus was presented with auditory stimulus. The abscissa is the frequency of the sound stimulus and ordinate is the rate the subject answered two visual stimuli. Error bars are the standard deviation of 5 subjects. **a** Auditory stimulation that presented 2000 Hz sound for the first sound and 2000–4000 Hz sound for the second sound (frequency raise). **b** Auditory stimulation that presented 2000–4000 Hz sound for the first sound and 2000 Hz sound for the second sound (frequency fall).

the possibility that auditory pathways processing those frequency information can be involved in the Sound-induced flash illusion.

References

Ladan Shams, Yukiyasu Kamitani, & Shinsuke Shimojo. (2000) What you see what you hear *Nature*, 408, 788.

Ladan Shams, Yukiyasu Kamitani, & Shinsuke Shimojo. (2002) Visual illusion induced by sound. *Cognitive Brain Research*, 14, 147-152.

Ladan Shams, Wei Ji Ma, & Ulrik Beierholm.(2005) Sound-induced flash illusion as an optimal percept. *Neuroreport*, 16, 1923-1927.

Orna Rosenthal, Shinsuke Shimojo, & Ladan Shams.(2009) Sound-Induced Flash Illusion is Resistant to Feedback Training. *Brain Topogr*, 21, 185-192.

NEUROPHYSIOLOGICAL METHOD FOR UNREVEALING THE AUDIO-VISUAL INTEGRATION IN MONGOLIAN GERVIL

Yuki Ito, Takafumi Furuyama, Shizuko Hiryu, and Kohta I. Kobayasi
Doshisha University, Kyotanabe, Kyoto, Japan

Abstract

This study tried to establish a methods to elucidate neural mechanism of the audio-visual integration in Mongolian gerbil. First, we evaluated a brain region related to the integration by evoked potential. Auditory and visual stimuli were presented either multi or unimodally. Waveform similarity between sum of each unimodal stimulus and audio-visual stimulus is different by location. Particularly, the region is laterally situated to primary visual cortex. This suggest that the region around primary visual cortex could be involved in the audio-visual integration. Second, we recorded visual mismatch negativity (vMMN). Most of the times, visual stimulus was presented from the same location (standard), and interval of presentation infrequently changed (deviant). As a result, difference (standard vs deviant) peaked around the 150–250 ms latency range, and is considered as vMMN. These results demonstrated that cortical mapping with vMMN recordings in Mongolian gerbil are tractable methods unrevealing physiological basis of the audio-visual integration.

We perceive the world by integrating information from different sensory modalities. Although the multimodal integration were thought to arise in the higher cortical processes, it is suggested that the integration also occurs in the lower cortex (primary and secondary sensory cortex) in recent studies. Mismatch negativity (MMN) is evoked by unexpected and rare stimuli in the environment, and is utilized by many studies to investigate lower cortical regions. While auditory evoked MMN (aMMN) is well known in both humans and rodents, visual MMN (vMMN) in rodents is still not well described. By observing vMMN in rodent, we could understand characteristics of its visual perception. Moreover, it can be used as physiological measure to reveal audio-visual integration, specifically for analyzing the cross modal influence on the visual sensation, in highly tractable model animal: rodent. The purpose of this study is to establish a methods to elucidate neural mechanism of the audio-visual integration in Mongolian gerbil.

Method

Experiment I

Subjects preparation. Three 20 to 48 week-old Mongolian gerbils (Meriones unguiculatus) were used for this experiment. Each subject was kept in individual cage, and all subjects were maintained on a 12-h light and 12-h dark schedule. All husbandry and experimental procedures were approved by the Animal Experiment Committee of Doshisha University.

The subjects were anesthetized with isoflurane during the experiment. Scalp fur was shaved off, and the skin over the cranium were removed. A head post (small metal rod with a bottom) was fixed to the top of the exposed cranium with instant acrylic glue and dental cement. The left cranium extending from Bregma and Lambda and 6 mm lateral to midline was removed, and cortical surface including visual cortex was exposed.

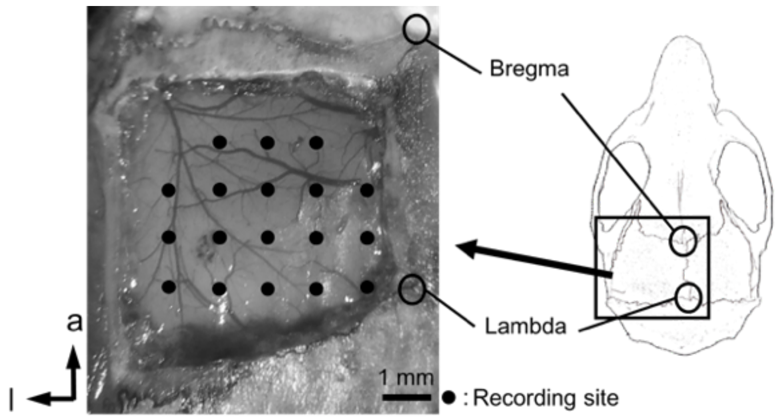


Fig. 1. Recording sites. Right figure is top view of the gerbil's skull. Left figure is top view of the surface and recording sites in the left hemisphere of gerbil. Closed circles are recording sites.

Recording procedure. A tungsten microelectrode (impedance 200–400 kΩ) was inserted into the visual cortex with a hydraulic micromanipulator. The recordings were obtained at intra-cortical depths (300 μm). The responses were amplified with a differential amplifier (DAM80, World Precision Instruments, USA) and filtered with a dual variable filter (0.1–5 kHz; VBF8, KEMO, UK). The recording responses were stored in a personal computer via an A/D converter (Micro 1401, Cambridge Electronic Design Limited, UK). The location of the electrode in primary visual cortex (V1) and surrounding V1 was determined at intervals of 1 mm based on Lambda (Fig. 1).

Stimuli. A white light-emitting diode (LED) was used as visual stimulus (V). The LED was positioned at 5 cm from right eye of a subject. We presented auditory stimuli (A) by using a sound card (UA-101, Roland Corporation, JPN) and loudspeaker (ES1, Tucker-Davis Technologies, USA). The loudspeaker was placed at 5 cm from right ear of the subject. The amplitude of the acoustic stimuli was calibrated with a microphone placed at the position of the head. We used 80 dB sound pressure level (SPL, re: 20 μ Pa) of broad-band noise (1–70 kHz). The durations of both visual and auditory stimuli were 3 ms with 1 ms rise/fall time.

The presentation of the auditory and visual stimulus alone was defined as A and V, respectively. In addition, both of A and V were presented simultaneously, called AV. Each type of stimuli was presented for 128 times.

Data analysis. The evoked potential to auditory stimulus was named as AEP, and the evoked responses to visual stimulus was called as VEP. The each type of evoked responses were averaged over the 128 repetition of stimuli, and the both latency and amplitude of responses to stimuli were quantified with the custom made programming (MATLAB). We created the waveforms summing of unimodal stimulus (A and V), called A + V, and compared the waveform similarity between A + V and AV in each recording sites.

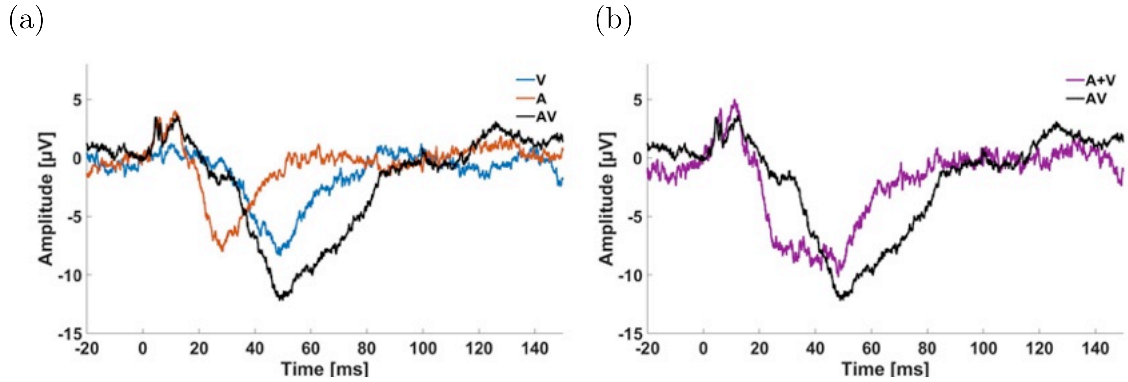


Fig. 2. (a) Waveforms when presented visual alone (V), auditory alone (A) and simultaneous stimulus (AV). Recording sites is 4 mm lateral and 2 mm anterior to Lambda. (b) Waveforms of sum evoked potential (A+V) and simultaneous evoked potential (AV).

Experiment II

In the oddball paradigm, a high-probability stimulus (standard) was presented with an interstimulus interval (ISI) of 800 ms and low-probability stimulus (deviant) was presented with ISI of 500 ms instead of ISI of standard. Deviant was presented for 60 times. The signal was amplified with a differential amplifier (DAM80, World Precision Instruments, USA) and filtered with a dual variable filter (0.1–300 Hz; VBF8, KEMO, UK). Recording site was the position of 2 mm lateral to Lambda. Other experimental systems are the same as Experiment I.

Results

Experiment I

Evoked potentials to visual, auditory, and bimodal auditory-visual stimuli at selected electrode sites are showed in Figure 2(a). ABR elicited by auditory stimuli was characterized within 10 ms from stimulus onset. VEP was characterized by a P1-N1 deflection at recording sites, after 40 ms respectively. The evoked potential, when auditory-visual stimulus presented, showed ABR and VEP as described for the alone stimuli. In Figure 2(b), differences between the sum evoked potential (A+V) and simultaneous evoked potential (AV) were observed. The waveform similarity between A+V and AV in each recording sites displayed in Figure 3. The position, 4 mm lateral and 2 mm anterior to Lambda, is lower than other positions in waveform similarity.

Experiment II

Figure 4 illustrates the evoked potentials to the standard before the deviant, those to the deviant and difference wave (the deviant minus the standard). The difference wave peaked around the 150–250 ms latency range.

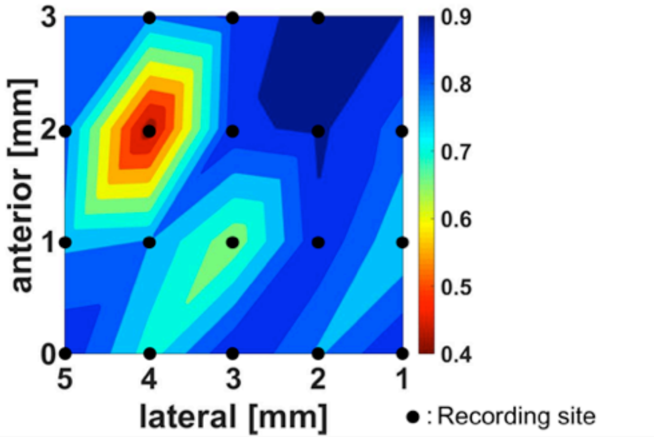


Fig. 3. Waveform similarity map between sum of each unimodal stimulus (A+V) and audio-visual stimulus (AV). Close circle is recording site.

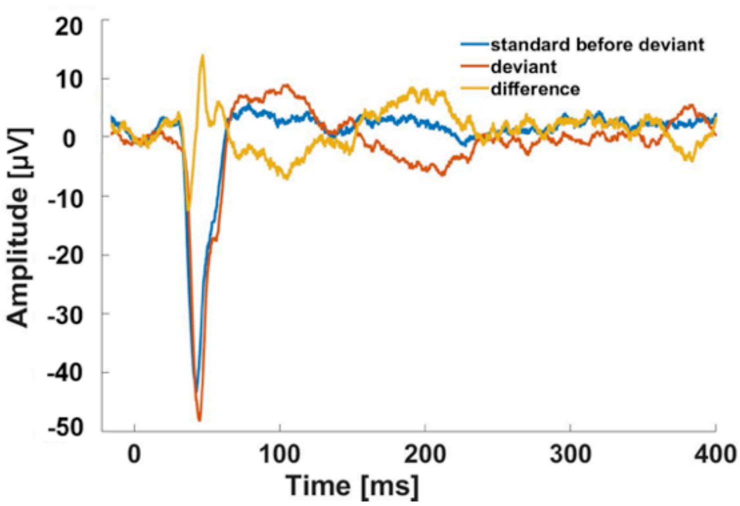


Fig. 4. Evoked potentials to the standard before deviant and deviant recorded in the oddball. Difference wave is the deviant minus the standard. The 0 ms indicates the onset of the stimulus.

Discussion

Waveform similarity (Figs. 2 and 3) suggests that in certain area multimodel stimulus could evoked diffarent cortical process which does not arise by unimodal stimulation. Therefore, it is considered that visual and auditory infomation integrate in the area. The lateral secondary visual cortex (V2L) is related to audio-visual integration in rats (Hirokawa et al., 2008). It is necessary to confirm where the region indicated by our data corresponds with. In primary auditory cortex (A1) of awake gerbils, auditory response decreased as the time lag between auditory and visual stimulus onset increased (Kobayasi et al., 2013). It is necessary that we examine the relationship of the time lag between two stimuli in the evoked potential in future study.

vMMN is described as a negative peak measured at the occipital electrodes between 150 and 350 ms after the simulus (Pazo-Alvarez et al., 2003; Czigler, 2007). The latency is comparable to our data, suggesting that diffarece wave we recoded is vMMN. These results demonstrated that vMMN recording in Mongolian gerbil is possible, and raise a posibility to invesitagate audio-visual integration using vMMN.

References

Czigler, I. (2007). Visual mismatch negativity: violation of non attended environmental regularities. *J. Psychophysiol.* 21, 224–230.

Hirokawa, J., Bosch, M., Sakata, S., Sakurai, Y., and Yamamori, T. (2008). Functional role of the secondary visual cortex in multisensory facilitation in rats. *Neuroscience*, 153, 1402–1417.

Kobayasi, K.I., Suwa, Y., Riquimaroux, H., 2013. Audiovisual integration in the primary auditory cortex of an awake rodent. *Neurosci. Lett.* 534, 24–29.

Pazo-Alvarez, P., Cadaveira, F., and Amenedo, E. (2003). MMN in the visual modality: a review. *Biol. Psychol.* 63, 199–236.

CROSS-MODAL EFFECT OF A SYNTHETIC SPEECH CONTINUUM ON THE JUDGMENT OF A VISUAL OBJECT'S SIZE

Sachi Itagaki[1], Hikari Oonishi[2], Shizuko Hiryu[1,2], and Kohta I. Kobayasi[1,2]
[1] *Graduate school of Life and Medical Science, Doshisha University*
1-3, Tataramiyakodani, Kyotanabe, Kyoto, Japan
[2] *Faculty of Life and Medical Science, Doshisha University*
1-3, Tataramiyakodani, Kyotanabe, Kyoto, Japan
<cygb1001@mail4.doshisha.ac.jp>

Abstract

Sound symbolism is a phenomenon that gives meaning and impression to sound itself. In a lot of the previous studies, stimuli were presented visually with alphabets, and subjects directly answered the impression of the sound. Therefore, it is not known which of basic acoustic features (e.g., duration and frequency), or perception of phoneme cause the sound symbolism. The purpose of this study is to evaluate the effect of phoneme perception on the phenomenon. We synthesized phoneme continuum as stimulus, and focused on sound symbolism in visual size perception. Voice sounds were assumed to have impression of bigger or smaller, according to previous researches. Subjects were all right-handed Japanese native speakers, and they didn't have knowledge about the sound symbolism and this experiment. As a result, effect of sound was small when perception of phoneme was difficult, suggesting that the phoneme perception is significantly related with sound symbolism.

Sound symbolism is an idea that sounds itself has certain impression (Ramachandran et al., 2001) (Kanero et al., 2014). For example, previous study reported that "p", "s" or "i" have smaller impression, and "b", "o" or "a" have larger impression (Newman, 1933) (Sapir, 1929). Our previous study confirmed that "p" and "i" have smaller impression and "b" and "o" have larger impression. The result corroborate that sound symbolism does occur even when the sound stimulus was presented aurally. In addition, high-frequency tone have smaller impression and low-frequency tone have larger impression (Gallace et al., 2006). Sound symbolism has been studied in vocal sound and pure tone, and following factors have been investigated as causes of the sound symbolism: frequency (or pitch of sound) and character of phoneme. Previous studies, however, have not systematically revealed the relative strength of each factors, let alone the interaction of these factors. Therefore, the purpose of this study is to clarify the influence of the perception of phoneme on sound symbolism. Furthermore, we compare the influences of consonants and vowels and examine the difference in their effects.

Materials and Methods

Subjects

Ten (6 females and 4 males; age range 21–24 years old) subjects participated in experiment 1, and 10 (8 females and 2 males; age range 21–23 years old) subjects participated in experiment 2. All subjects were native Japanese speaker. They have neither knowledge of phonetic symbolism nor these experiments.

Visual Stimulus

This study examined the phonetic symbolism in magnitude judgment of visual stimulus. Different size of gray circles like a doughnut were presented on black background LCD display as a visual stimulus. Standard had outer circle of 300 pixels and inner circle of 280 pixels. Target size was either smaller or bigger than the standard by ± 40% in diameter. These visual stimuli were same in experiment 1 and 2. Standard was presented for 800 ms and target was presented for 500 ms. Screen size was 1024 × 768 pixels. A red cross (34 pixels) was always presented as a fixation point at the center of the screen.

Sound Stimulus

The morphing sounds were presented as sound stimuli to examine influence that phoneme perception give to sound symbolism. A speech analysis and synthesis software, "STRAIGHT" (Kawahara et al., 1999), was used to generate 10-step audio-morphed continua between pairs of naturally recorded speech (/pi/–/ba/ and /i/–/a/ speech syllables). For instance, for each pair, we generated 10 intermediate tokens as 10% acoustic steps from 0% (highly similar to "pi" or "i") through to 100% (highly similar to "ba" or "a"). In experiment 1, [pi : ba] = [100% : 0%], [90% : 10%], [80% : 20%], [70% : 30%], [60% : 40%], [50% : 50%], [40% : 60%], [30% : 70%], [20% : 80%], [10% : 90%], [100% : 0%]. And, in experiment 2, [i : a] = [100% : 0%], [90% : 10%], [80% : 20%], [70% : 30%], [60% : 40%], [50% : 50%], [40% : 60%], [30% : 70%], [20% : 80%], [10% : 90%], [100% : 0%]. Around 50% sounds were likely to be perceptually the most ambiguous and may, for example, be interpreted as "ba" or "pi" depending on the listener.

Experimental Paradigm

Subjects were required to answer visual size difference between the standard and the target stimulus. We used the same experiment paradigm for experiment 1 and experiment 2. A trial began with 2000 ms rest, then standard stimulus of 800 ms was presented, and after 700 ms inter-stimulus interval (ISI). Next, target stimulus of 500 ms was presented followed by a response period of 2000 ms (Fig.1). During rest and response periods, the screen was black. After the subject answered to the task by pressing the button, next trial automatically began. They pressed the left and right button using the index finger and ring finger, respectively. Which button (right or left) represented which answer (bigger or smaller) was varied between subjects. There are 2 (1 Standard × 2 Target; visual) and11 (sound) combination of each stimulus, respectively; taken together, the combinations of the overall stimulus were 22 types. The entire stimulus type was randomly order to create one block (equals 22 trials). One session had total 3 blocks. In both experiment 1 and experiment 2, each subject carried out 2 sessions (equals 132 trials). There was a 5 minutes break between sessions.

Results and Discussion

We defined the congruent and incongruent conditions as follows and analyzed the data accordingly. The congruent condition occurred when the target stimulus was consistent with the impression made by the sound (i.e., the larger target was presented with "ba"/"a" or the smaller with "pi"/"i"). The incongruent condition occurred when the target stimulus

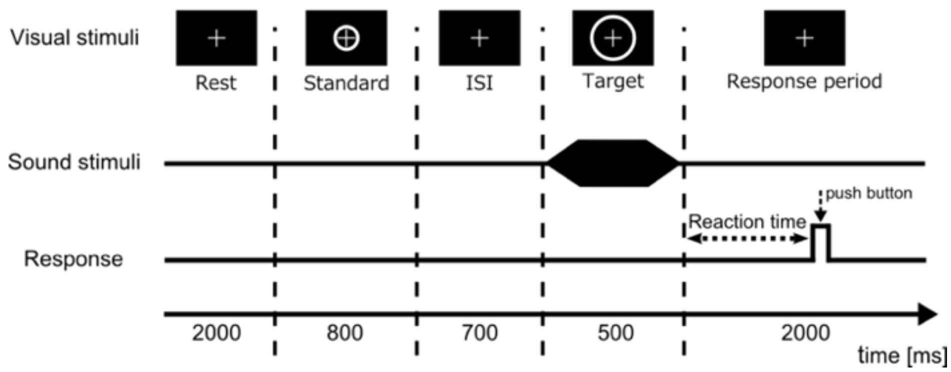


Fig. 1. Presentation of visual stimulation and sound stimulation and flow chart of action response. A red cross was always presented as a fixation point.

was inconsistent with the impression of the sound (i.e., the larger target was presented with “pi”/“i” or the smaller one with “ba”/“a”).

Experiment 1

The differences in Z score of RT between congruent and incongruent conditions that sound stimuli were clearly “ba” and “pi” are depicted in Figure 2A. The mean Z score of RT under the incongruent condition was longer than that under the congruent condition. The differences in Z score of RT between congruent and incongruent conditions for all sound stimuli are depicted in Figure 2B. In the case that component of “ba” was less than 50%, Z score of RT under smaller circle change was shorter. On the other hand, in the case that component of “ba” was 50% or more, Z score of RT under larger circle change was shorter.

Experiment 2

The differences in Z score of RT between congruent and incongruent conditions that sound stimuli were clearly “a” and “i” are depicted in figure 3A. The mean Z score of RT under the incongruent condition was longer than that under the congruent condition, and the difference was significant ($t = -2.48, p < 0.05$). The differences in Z score of RT between congruent and incongruent conditions for all sound stimuli are depicted in Figure 3B. In the case that component of “a” was 50% or less, Z score of RT under smaller circle change was shorter. On the other hand, in the case that component of “a” was more than 50%, Z score of RT under larger circle change was shorter.

Figures 2A and 3A show that the Z-score of reaction time in congruent condition are shorter than that in incongruent condition in both experiment 1 and experiment 2. These results suggest that sound symbolism did occur. In other words, “ba” and “a” have larger impression and “pi” and “i” have smaller impression. Additionally, it was suggested that sound symbolism occurs more in the sound stimulation of only vowel than in the combination of consonant and vowel, suggesting that influence by vowel is larger than consonance. In figures 2B and 3B, sound symbolism in relationship between sound and magnitude of objects was changed by the perception of phoneme. The change is

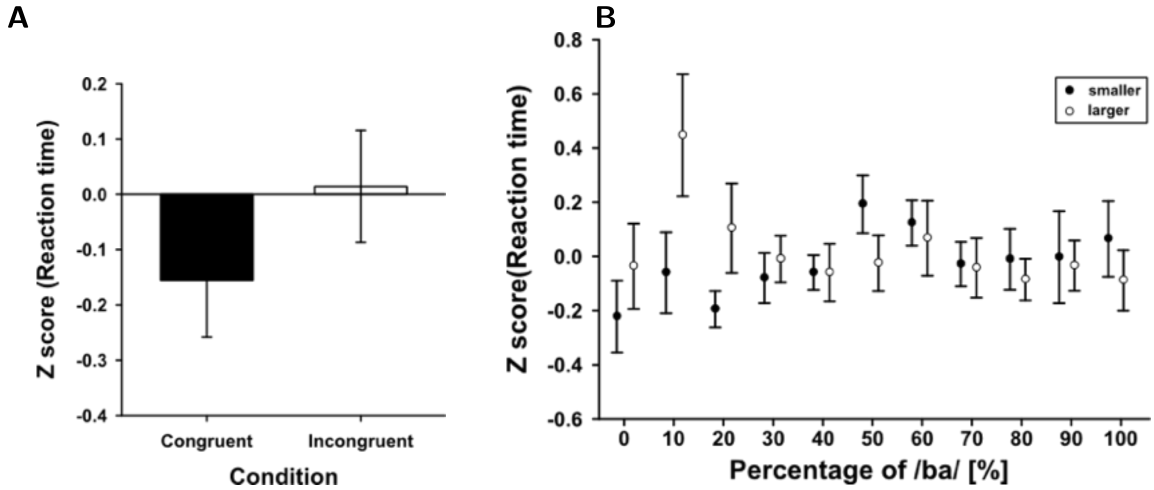


Fig. 2. (A) Average reaction time of all circle size condition depicted as Z score. Error bars represent the standard deviation of mean. A black bar is congruent condition and a white bar is incongruent condition. (B) Average reaction time of each sound condition depicted as Z score. Error bars represent the standard deviation of mean. Black circles are smaller condition, and white circles are larger condition.

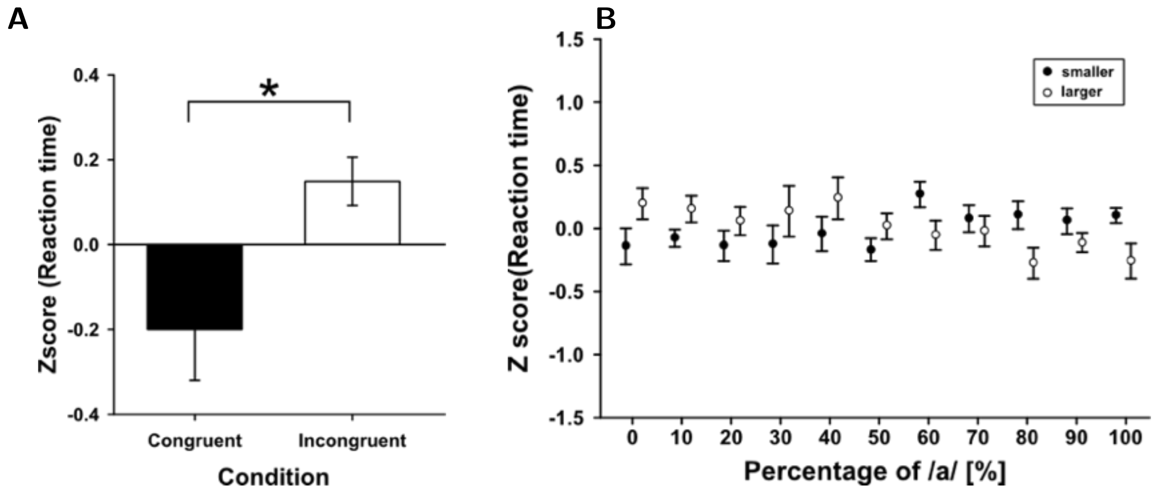


Fig. 3. (A) Average reaction time of all circle size condition depicted as Z score. Error bars represent the standard deviation of mean. Black bar is congruent condition and white bar is incongruent condition. Reaction time in congruent condition is shorter than in incongruent ($t = -2.48, p < 0.05$). (B) Average reaction time of each sound condition depicted as Z score. Error bars represent the standard deviation of mean. Black circles are smaller condition, and white circles are larger condition.

thought to be greater when the target circle changes smaller than standard circle.

References

Gallace A., and Spence C., "Multiseneory Synesthetic Interactions in the Speeded Classification of Visual Size", *Percept. Psychophys.*, 68(7), 1191–1203 (2006).

Kanero J., Imai M., Okuda J., Okada H., and Matsuda T., "How Sound Symbolism Is Processed in the Brain: A Study on Japanese Mimetic Words", *Plos ONE*, 9, e97905 (2014).

Newman S. S., "Further Experiments in Sound Symbolism", *Am. J. Psychology*, 45, 53–75 (1933).

Ramachandran V. S. and Hubbard E. M., "Synesthesia—A Window Into Perception, Thought and Language", *J Consciousness Stud*, 8, 3–34 (2001).

Sapir E., "A study in phonetic symbolism", *J. Exp. Psychol.*, 12, 225–239, (1929).

DOES FACE COLOR FACILITATE TO JUDGE THE FACIAL EXPRESSION?

Fumiyo Takahashi and Yasuhiro Kawabata
Department of Psychology, Hokkaido University
N10W7, Kita ward, Sapporo, Hokkaido 0600810, Japan
<t-fumiyo@let.hokudai.ac.jp>

Abstract

The hypothesis that humans' trichromatism has developed to discriminate skin colors suggests that face color is important as a part of a facial expression. Therefore, we investigate the effects of face color on the judgment of facial expressions. We used schematic faces as stimuli to control the degree of intensity of facial expressions. Each facial expression had five color variations. Four of those corresponded to emotions—red, blue, yellow, white—while one did not correspond to any emotions—green. Each expression with each of the five colors was presented one by one. Participants answered whether or not a face was emotional, and chose, which emotion they read, from alternatives. The results revealed that the correct responses were increased in red of anger, and blue of sadness. Moreover, the participants' reaction times were shorter in red of anger and blue of sadness. It suggests that face colors corresponding to emotion facilitate to judge facial expressions.

It is considered that color plays an important role in visual cognition. The effects of color were reported only in the diagnostic tasks for scene gist and for object and recognition with typical color (Yip, W. A. & Sinha, P., 2002; Russell, R., et al., 2006; Castelhano, M. S., & Henderson, J. M., 2008; Bindemann, M. & Burton, A.M., 2009; Lloyd-Jones, T. J., & Nakabayashi, K., 2009; Elder, J. H. & Velisavljevic, L., 2009). Besides, using computational models, Changizi, Zhang, and Shimojo (2006) showed, in terms of evolution, that the trichromatopsia of primates, whose faces and breeches are exposed, was developed in order to show and to be grasped by others, the color changes due to the circulation of blood fluid on face or on breech. Accordingly, it is suggested that typical color for facial expression could exist and it could work well in the judgment of facial expression, which is widely considered important as a social signal. The present research investigated the role of face colors in the judgment of facial expression. We hypothesized that face colors or typical colors for facial expressions are significant factors in identifying facial expressions. Our research questions are as follows, 1) Does typical color of each emotion facilitate performance on the cognitive task? 2) The harder to judge emotion, the more color effects? Since the previous study showed the color effect appeared when the target is hardly identified (Yip, W. A. & Sinha, P., 2002; Castelhano, M. S., & Henderson, J. M., 2008). We presumed that color effect will appear in difficult trial such as weak and ambiguous facial expression. We used two typical colors such as red to *anger* and blue to *sadness,* in which Takahashi and Kawabata (2017) found a clear tendency in the evaluation of color association with schematic facial expressions.

Method

The stimuli were presented in 24"LCD displays by the SuperLab pro 4.5. We used the stimuli that were used in our previous research to evaluate the threshold of facial expres-

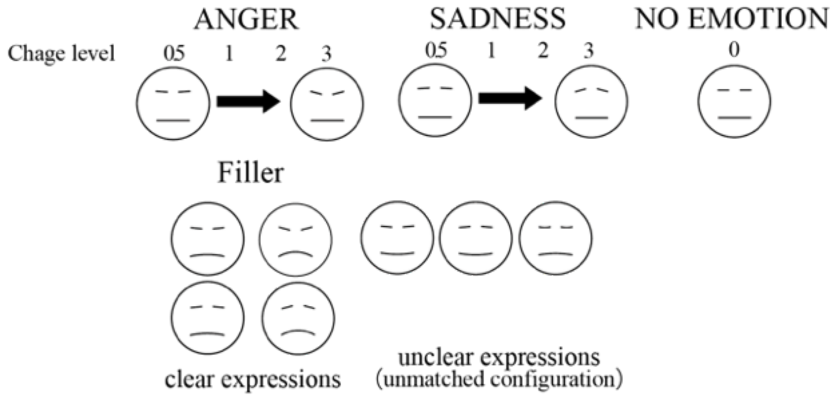


Fig. 1. The variations of facial expression. The intensity of each emotional expression on the schematic face is regulated by changing the tilt angle of the eyes. The change level corresponded to the tilt degree.

sions (Takahashi, F., & Kawabata, Y., 2011). Two emotional facial expressions, *anger* and *sadness* were used as target stimuli. The intensity of facial expressions was regulated with four steps by changing the tilt angle of eyes. The filler stimuli were with obvious expressions to targeted emotions (Takahashi, F., & Kawabata, Y., 2011) and other complex expressions. Each facial expression had five color variations, red, blue, yellow, green, and white. The color variations were selected based on the results of our previous study (Takahashi, F. & Kawabata, Y., 2017). Four of those corresponded to emotions—red for *anger,* blue for *sadness,* yellow for joy, and white for no emotion—, while green did not correspond to any emotions. The patterns of stimuli were eighty (Fig. 1). 80 participants data were performed for analysis. The participants were Japanese students and staffs in Hokkaido University. All had normal or corrected-to-normal visual acuity and normal color vision. There were 36 males ($M = 22.8, SD = 5.43$) and 44 females

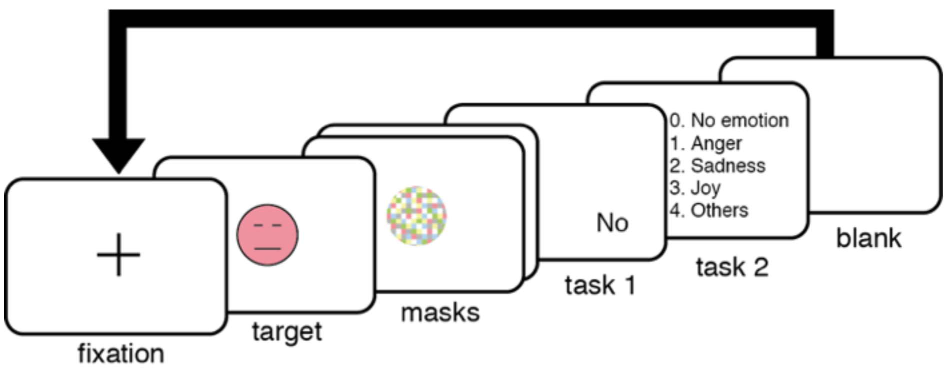


Fig. 2. The procedure of the trial. It started with a fixation for 1000 ms and followed by a target face for 100 ms. Subsequently, two mask stimuli were presented and presented "Yes" and "No" corresponded to the side of answer key. After pressing the key, the alternatives were presented to make participants choose. After choosing the number by pressing a numerous key, followed by a blank screen and returned to the start of the trial.

($M = 21.0, SD = 4.88$).

One block included 80 trials and four blocks were repeated and measured. The participants practiced 10 trials after experimenter's instruction before the experiment. In each trial, firstly, the fixation was presented at the center for 1000 ms. Subsequently, a schematic face was presented at the center for 100 ms on white background. Two kinds of mask stimuli followed for 50 ms each (Fig. 2). After this presentation, participants were required to respond to whether the face was emotional or not by pressing a key as quickly and accurately as possible (task 1). Followed by alternatives were presented and participants chose, which emotion do they had read on the face by pressing a numeric keypad, from them (task 2). The position of response key in task 1 is counter balanced between participants.

Results

In task 1, participants judged whether or not a face was emotional, and in task 2, participants chose, which emotion they read, from alternatives. Fig. 3 shows the mean of the reaction time (RT) of task 1 and mean of the correct response (CR) of task 2 by each emotion. The abbreviations of x-axis are as follows, ANG, *anger;* SAD, *sadness;* NEU, neutral as *no emotion.* From the result, for *anger,* the RT is the shortest and the CR is the highest. For *sadness,* the results are vice versa. These results suggest that expression of *sadness* is relatively difficult to identify the emotion. For *anger* and *sadness,* two-way repeated-measure ANOVA was performed ($N = 80$). The factors were change level with four levels and color with five levels. For *no emotion,* one-way repeated-measure ANOVA was performed on color. An alpha level of .05 was used for all statistical tests.

Fig. 4 shows the results of *anger.* The first line shows the RT of task 1, and the left side shows those by each color. This indicates that RT in red is shorter than in the other colors. The right graph shows the reaction time by the change level, and the each line depicts the result by each color. Although there is no difference by color in change level 0.5, the difference was the maximized in change level 1. The second line shows the CR of task 2. This indicates that CR in red is higher than the other colors. The right

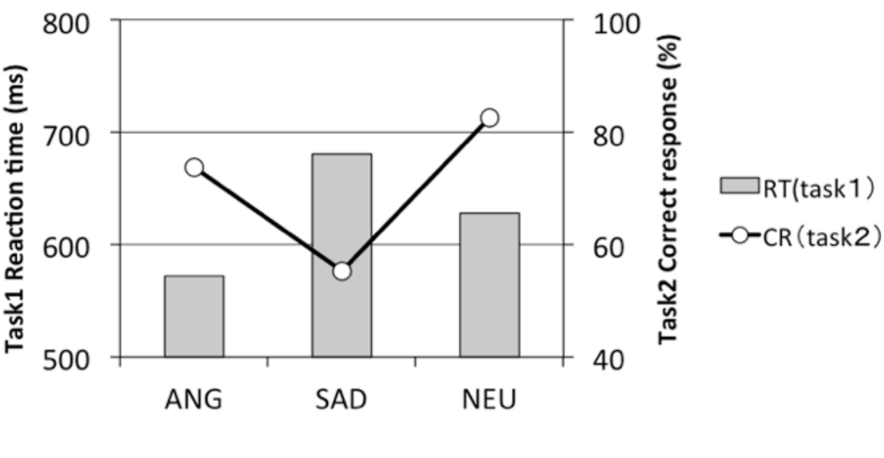


Fig. 3. Mean of the reaction time of task 1 and mean of the correct response of task 2 by each emotion. The abbreviations of x-axis are as follows, ANG, *anger;* SAD, *sadness;* NEU, neutral as *no emotion.*

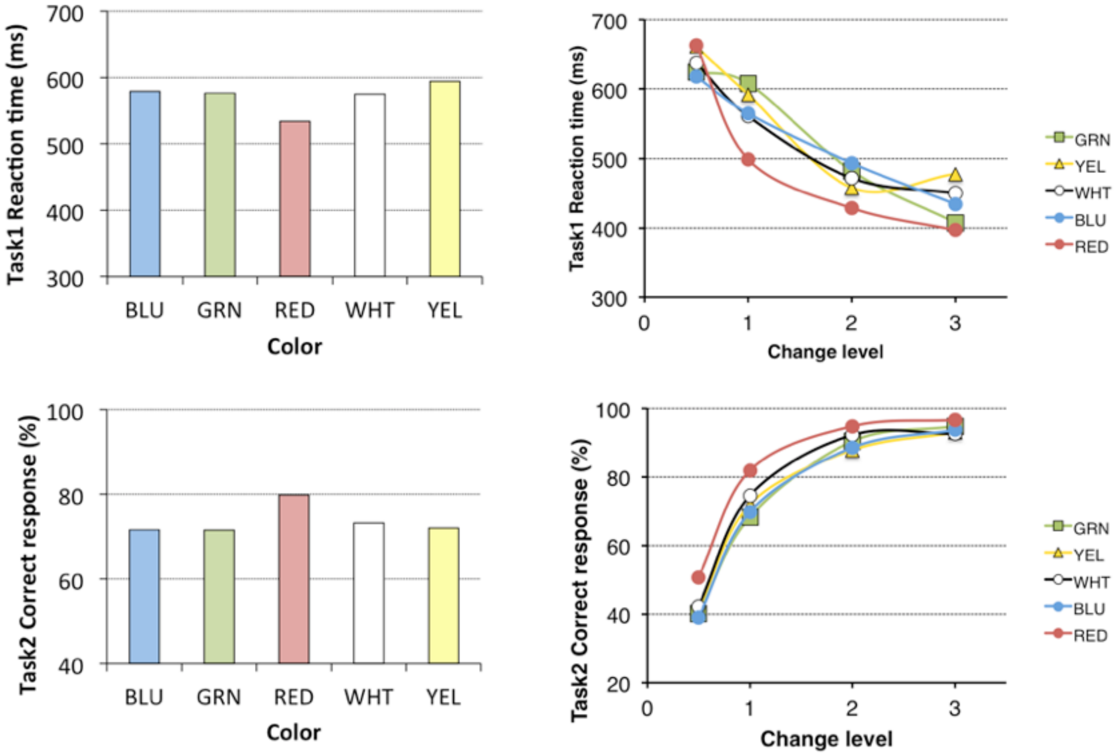


Fig. 4. The results of *anger.* Upper line shows the RT of task 1, and lower line shows the CR of task 2. The left side shows the mean of the RT of task 1 and mean of the CR of task 2 by each color. The abbreviations on x-axis are as follows, BLU, blue; GRN, green; RED, red; WHT, white; YEL, yellow. The right side shows the mean of the RT of task 1 and mean of the CR of task 2 by the change level. The each line of left graphs depicts the result by each color.

side graph indicates that the CRs in red are higher though the differences are different by change levels.

Fig. 5 shows the results of *sadness.* The first line shows the RT of task 1, and the left side shows those by each color. The first line indicates that RT in blue is shorter than the other colors. In the right graph, RT in blue is shorter in change level 2 and 3. The second line indicates that CR in blue is higher than the other colors at the all of the change level. These results indicate that color effect might depend on the easiness of identifying the emotion. Moreover, Fig. 6 shows the results of no emotion. The highest performance is at white corresponding to *no emotion.* This result is consistent with *anger* and *sadness* in the CR in task 2. However, the RT is not the shortest and rather red and blue were longer.

Discussion

Does Typical Color of Each Emotion Facilitate Performance on the Cognitive Task?

The facilitation effect appeared in all emotions when the color of target faces corresponded to their emotion. Especially, these color effects in *anger* and in *sadness* tended to be

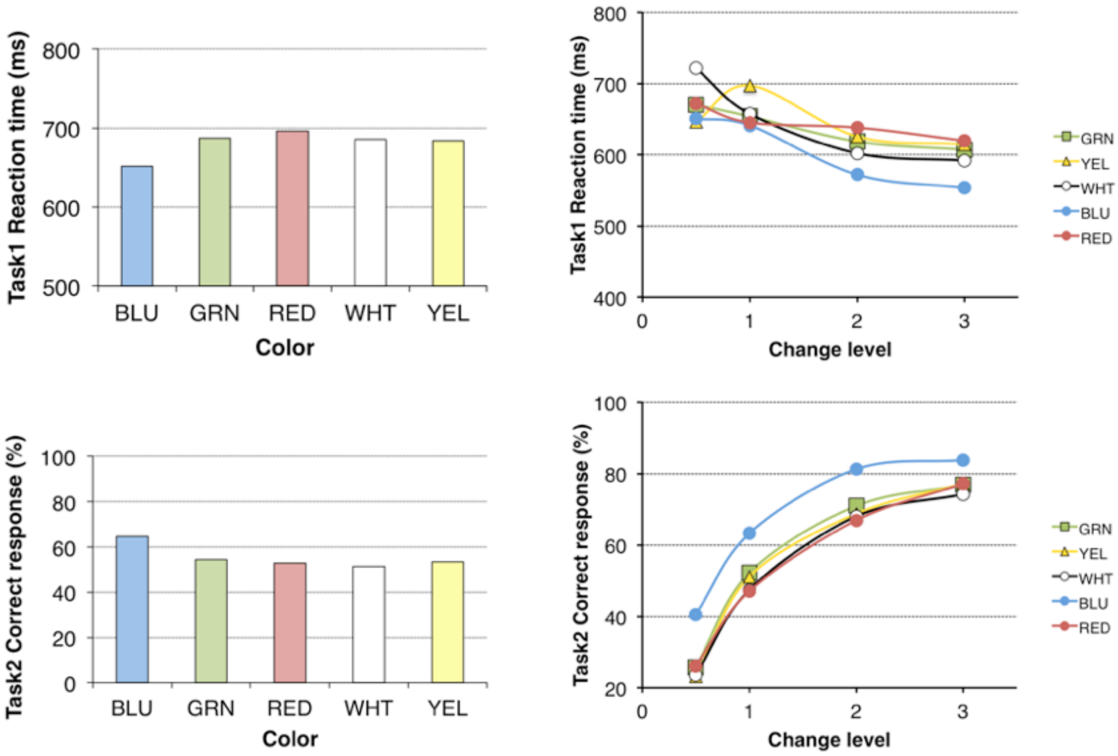


Fig. 5. The results of *sadness.* Upper line shows the RT of task 1, and lower line shows the CR of task 2. The left side shows the mean of the RT of task 1 and mean of the CR of task 2 by each color. The right side shows the mean of those by the change level. The each line of left graphs depicts the result by each color.

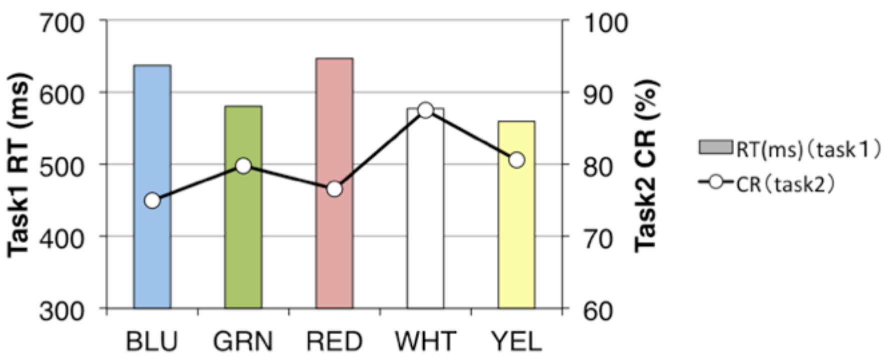


Fig. 6. The results of *no emotion.* Mean of the reaction time of task 1 and mean of the correct response of task 2 by each emotion.

salient in the specific intensity of facial expressions. The results of CR and of RT showed different tendencies among emotions. Accordingly, it is suggested that the differences could be caused by different systems of identifying or detecting among emotions, or by different mechanisms of emotional biological readiness. Presumably, these might cause the different tendencies among emotions. The RT of *no emotion* was delayed in red and blue. Some inhibition effects of untypical color, which corresponded to a specific emotion, might affect the judgment for emotion.

The Harder to Judge Emotion, the More Color Effects?

The color effects appeared saliently in some intensity of facial expressions. For reaction time, *anger* was faster in the change level 1, 2 and 3, and *sadness* was faster in the change level 2 and 3. These results suggest that there were appropriate conditions that color affects to judge the emotion. Hence, our hypothesis that color effect appears in difficult trial such as faint or ambiguous facial expression was supported. For the correct response of *no emotion* was lower in blue and red significantly. These suggest cognitive conflict, and the typical color of other facial expressions could cause inhibition effect such as red to *anger* and blue to *sadness.*

References

Bindemann, M., & Burton, A. M. (2009). The role of color in human face detection. *Cognitive Science,* 33(6), 1144–56.

Castelhano, M. S., & Henderson, J. M. (2008). The influence of color on the perception of scene gist. *Journal of Experimental Psychology: Human Perception and Performance,* 34(3), 660–75.

Changizi, A. M., Zhang, Q., & Shimojo, S. (2006). Bare skin, blood and the evolution of primate colour vision. *Biology Letters,* 2(2), 189–195.

Elder, J. H., & Velisavljevic, L. (2009). Cue dynamics underlying rapid detection of animals in natural scenes. *Journal of Vision,* 9, 1–20.

Lloyd-Jones, T. J., & Nakabayashi, K. (2009). Independent effects of colour on object identification and memory. *The Quarterly Journal of Experimental Psychology,* 62, (2), 310–322.

Russell, R., Sinha, P., Biederman, I., & Nederhouser, M. (2006). Is pigmentation important for face recognition? Evidence from contrast negation. *Perception,* 35(6), 749–759.

Takahashi, F., & Kawabata, Y. (2011). The color effect on identifying the emotions of facial expression: Facial expression recognition threshold for shape change of facial parts, the 75th Annual Convention of the Japanese Psychological Association, Nihon University, Tokyo (Japan), *Proceedings,* 702.

Takahashi, F., & Kawabata, Y. (2017). The association between colors and emotions for emotional words and facial expressions. *Color Research and Application* (in press).

Yip, A. W., & Sinha, P. (2002). Contribution of color to face recognition. *Perception,* 31(8), 995–1003.

STUDY ON THE CRITERIA FOR ARTISTIC ACTIVITIES BY PEOPLE WITH DISABILITIES: VERIFICATION OF THE CRITERIA LISTS BY SUBJECTIVE EVALUATION TEST FOR THE EXPERTS

Tsukasa Muraya[1], Takeharu Seno[2], and Yasuyuki Hirai[3]
[1]*Graduate School of Design, Kyushu University, Fukuoka 815-8540, Japan*
[2]*Faculty of Design/Research Center for Applied Perceptual Science*
Kyushu University, Fukuoka 815-8540, Japan
[3]*Faculty of Design, Kyushu University, Fukuoka 815-8540, Japan*

Recently, artistic activities by people with disabilities have evolved beyond the boundaries between art and welfare. However, holistic evaluation method for those activities have not been created yet. This study focused on the artistic activities process by people with intellectual disabilities. Firstly, this study collected some viewpoints and approaches for the evaluation by comparing five literatures written by five most important experts of artistic activities by people with disabilities. Two criteria were created based on it. However, whether these criteria really can be used and valid in the various scenes of artistic activities was not examined. Thus, secondly, this study challenged to verify the importance of these criteria lists. Subjective evaluation test about the list (6-point-rating scales) was used for the same five experts. As a result, this study identified correlations and differences among them with numeric values. These objective values can newly suggest us the validity of our criteria.

Part X
Poster Session 2

CAN THE DELAY TIME OF SUBJECT'S RESPONSE BE A TRUE COMPONENT OF THE THEORY OF HUMAN INTERMITTENT CONTROL?

Takashi Suzuki and Ihor Lubashevsky
University of Aizu
Ikki-machi, Aizu-Wakamatsu, Fukushima 965-8580, Japan
IL: <i-lubash@u-aizu.ac.jp>

Abstract

We report the results of our experiments on human response delay in balancing virtual pendulums with overdamped dynamics. The overdamping eliminates the effect of inertia, thereby, reduces the dimensionality of the system under control, and enables us to measure the delay time directly in the stream of subject's actions. Based on these results we demonstrate that the delay time of human response, at least under the analyzed conditions, is an essentially random variable characterized by (i) a wide distribution, (ii) weak correlations in the sequence of subject's actions, and (iii) the substantial dependence of its probabilistic properties on the particular details of the balancing. Besides, we raise a doubt about the universality of the classical approach to modeling human intermittent control based on delay-differential equations because their generalization going beyond the fixed delay time approximation becomes extremely overcomplicated and, as a result, intractable.

Human Intermittent Control and Subject's Response Delay

Although the intermittency of human control over unstable systems is recognized for decades its mathematical description is up to now a challenging problem. Recently models based on the concept of event-driven control activation have become much employed (Milton et al., 2009; Gawthrop et al., 2011; Kowalczyk et al., 2012; Milton, 2013). These models take into account that humans cannot detect small variations in the state of a system under control, thereby, the control is activated by subjects when the system deviates rather far from the desired equilibrium and halted after being returned to its proximity.

Within this approach the threshold mechanism of control activation with the fixed delay time of subject's response is widely accepted as the pivot point in constructing specific mathematical models for such human actions. For illustration, let us consider a typical model for the balancing of inverted pendulum. It assumes the pendulum to be governed by the following equation written for the angle θ between the pendulum at the current moment of time t and the upright position:

$$\frac{d^2\theta}{dt^2} + k\frac{d\theta}{dt} - \omega^2\theta = \mathcal{R}\left[\theta|_{t-\tau}, \left.\frac{d\theta}{dt}\right|_{t-\tau}, \left.\frac{d^2\theta}{dt^2}\right|_{t-\tau}, \xi(t)\right]\cdot \mathrm{H}\Big(|\theta| - \theta_c\Big), \qquad (1)$$

for a review see, e.g., Milton (2013); Milton et al. (2015); Insperger et al. (2015). The left-hand side of Eq. (1) represents the effects of pendulum inertia, gravity, and viscous friction. Its right-hand side describes subject's response to the pendulum deviation from the upright position θ, the angular velocity $d\theta/dt$ and, maybe, acceleration $d^2\theta/dt^2$ the subject perceives with some delay time τ. The right-hand side of Eq. (1) also contains white noise $\xi(t)$ allowing for random factors in human actions generally depending on the system state (e.g., Cabrera and Milton, 2002; Cabrera et al., 2004; Hosaka et al., 2006). The Heaviside step function $\mathrm{H}(\ldots)$ is due to the perception threshold mechanism.

It should be noted that there are models—mainly dealing with discrete representation of system dynamics—considering possible ways of introducing random variations into the delay time and temporal non-locality into the description of human intermittent control (e.g., Ohira, 2007a,b, 2009). Besides, there have been developed models based on equations similar to Eq. (1) for describing human behavior in the balancing process admitting also the effects of anticipation (for a review see, e.g., Ohira, 2009; Gawthrop et al., 2011; Insperger and Milton, 2014). These models assume the information about the system state at time moments $t' < t - \tau$ to be known and use the internal models to "calculate" the system states in the time interval $(t - \tau, t)$.

Recently (Zgonnikov et al., 2014) we have proposed an alternative mechanism of transitions between passive and active phases of human intermittent control—noise-induced control activation being intrinsically stochastic, noise-driven. It suggests that control activation stems from stochastic interplay between subject's need to keep the controlled system near the desired state, on one hand, and the tendency to postpone interrupting the system dynamics, on the other hand. This mechanism falls outside the scope of threshold approximation in describing control activation and if admits more complex properties of human response.

Results of Experiments on Balancing Overdamped Pendulum

In the present work we report the results of our experiments on pendulum balancing within a setup similar to one used in our previous experiments (Zgonnikov et al., 2014) and raise a doubt about, at least, the universality of the formalism based on delay-differential equations in describing such human actions.

The used experiment setup is illustrated in Fig. 1. The pendulum dynamics is simulated by numerically solving the equation

$$\tau_\theta \frac{d\theta}{dt} = \sin\theta - \frac{\tau_\theta}{l}\eta\,\vartheta\cos\theta\,, \quad (2)$$

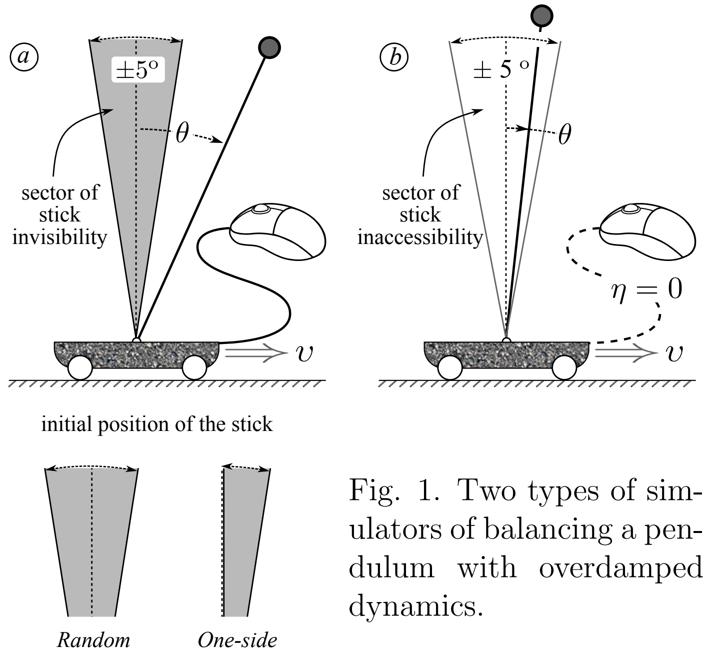


Fig. 1. Two types of simulators of balancing a pendulum with overdamped dynamics.

where θ is the angular deviation of the stick (pendulum) from the vertical position and ϑ is the cart velocity. The parameter τ_θ determines the time scale of the stick motion, the stick length l determines the characteristic magnitude of the cart displacements required for keeping the stick upright. The cart position is controlled by subjects via the computer mouse. The cofactor η determines the feasibility of stick motion correction via the mouse movements; if $\eta = 1$ the stick is accessible and it is not accessible when $\eta = 0$. In these experiments:

(a) the stick becomes invisible within the sector $|\theta| < \theta_c$ and the two types of initial stick position were used such that the side on which the stick becomes visible is unpredictable ("Random") or predictable ("One-side");

(b) the stick remains visible but becomes unaccessible within the sector $|\theta| < \theta_c$ after the current active phase of its correction is finished.

The critical value of the stick angle θ_c was set equal to $\theta_c = 5^o$ in these experiments.

After the stick becomes visible or accessible for control a subject can try to return it into a small neighborhood of the upright position via moving the mouse. When this action becomes successful, the subject has no need to continue the mouse movement, which is the direct consequence of the stick overdamped dynamics. In this case when the cart velocity drops below a certain threshold, the control is halted, the stick is returned to the initial position, and the next trial of stick balancing begins. Figure 2 illustrates the measurement of the delay time of subject's response.

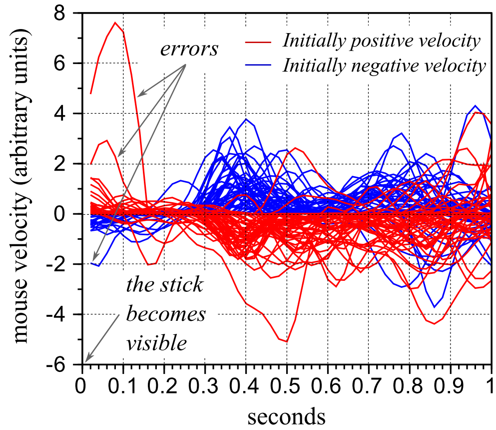
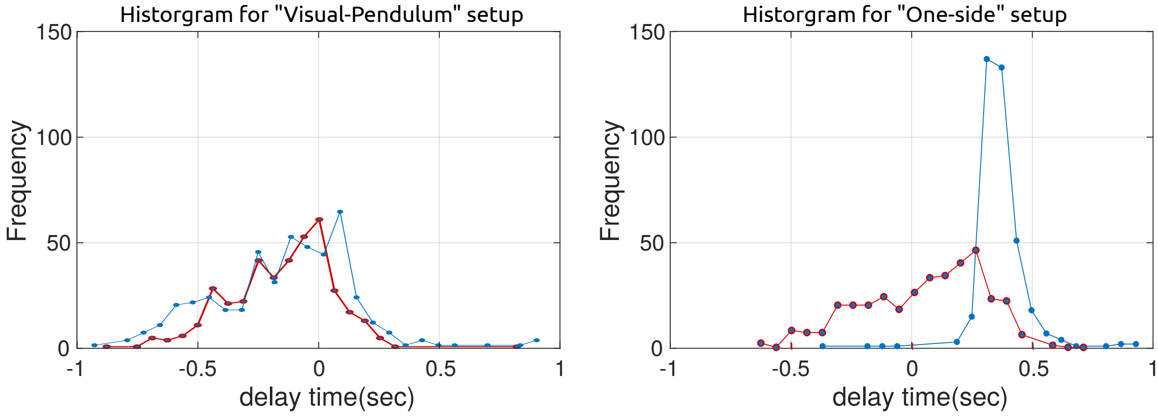


Fig. 2. Qualitative explanation of the measurement of the subject's response delay, e.g., in the experiments using the simulator shown in Fig. 1(a).

In the reported experiments 10 subjects were involved and for each subject the order of the conducted experiments was the balancing of the "visible" stick (Fig. 1b), followed by the initially invisible stick with the predicted side of its appearance (Fig. 1a, "One-side" version), and, finally, the initially invisible stick appearing randomly from both the sides (Fig. 1a, "Random" version).

The collected data were analyzed using the distribution of delay time constructed individually for all the subjects and the cumulative ones, the correlation functions of delay time in the serious of control events, and the corresponding time patterns. Figure 3 illustrates the found histograms. The obtained results together with the previous ones (Suzuki et al., 2015) argue for the following properties of the human response delay, at least in balancing pendulums with overdamped dynamics.

- The human response delay time is a random variable distributed inside a wide interval. The lower boundary of this interval can be less than 50 ms—typical response time determined by human physiology; its upper boundary is about 500–600 ms—the typical value for human response delay during complex balancing tasks.
- The response delay time is characterized by weak correlations in the stream of subject's actions in stick balancing and for some subjects the corresponding patterns in this series exhibit trends caused by subject's learning of stick balancing.
- In the experiments with the "visible" stick subject's delay time is affected substantially by the anticipation such that the distribution of delay time contains a substantial part corresponding to the negative values of delay time.
- In the experiments with the "invisible" stick for some subjects the histograms can exhibit a strong dependence on the predictability of stick appearance. Namely, for these subjects the distributions of delay time in the case of the "One-side" set-up are different from that of the "Random" set-up and can contain tails corresponding to negative values of delay time. The latter is caused by the effect of anticipation.
- Human individuality, on one hand, can be significant as noted in the previous item. On the other hand, the other characteristic properties of the delay time are rather similar including their magnitude for all the subjects.



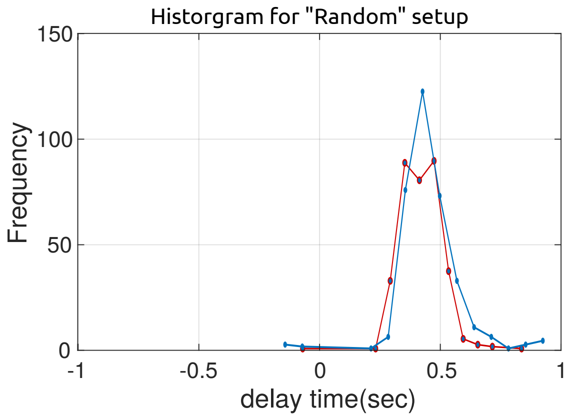


Fig. 3. Delay time histograms for two subjects with different behavior exhibited in balancing the pendulums within the "One-side" setup.

It should be emphasized that the delay time measured in our experiments is a result of cumulative contribution of several mechanism, including visual perception of stick motion, its deliberate analysis with possible anticipation, and the propagation of neural stimuli to muscles.

Conclusion

The found properties of subject's response have been found also in other investigations and individually admit a description generalizing the technique of delay-differential equations (see, e.g., Lakshmanan and Senthilkumar, 2010, as well as citations noted above). However, in our experiments *all* of them are met in dealing with actually *one* system. Therefore for describing human intermittent control in the given case within the formalism of delay-differential equations its generalization must be implemented via *several different* steps. Each of these steps individually endows the model with essential complexity and its resulting form seems to be overcomplicated. The latter, from our point of view, makes such models just intractable. It prompts us to suppose that there should be another formalism allowing for all of these features as its *inherent* properties. In particular, this formalism has to take into account the fundamental uncertainties in the human perception as well as the mental evaluation of system spatial position and instants of time. As a plausible candidate, the desired formalism can deal with space-time clouds rather than spatial coordinates of system position and point-like instants of time (Lubashevsky, 2017).

References

Cabrera, J. L., Bormann, R., Eurich, C., Ohira, T. and Milton, J.: 2004, State-Dependent Noise and Human Balance Control, *Fluctuation and Noise Letters* **4**(1), L107–L117.

Cabrera, J. L. and Milton, J. G.: 2002, On-Off Intermittency in a Human Balancing Task, *Physical Review Letters* **89**(15), 158702(1–4).

Gawthrop, P., Loram, I., Lakie, M. and Gollee, H.: 2011, Intermittent control: a computational theory of human control, *Biological Cybernetics* **104**(1-2), 31–51.

Hosaka, T., Ohira, T., Luciani, C., Cabrera, J. and Milton, J.: 2006, Balancing with noise and delay, *Progress of Theoretical Physics Supplement* **161**, 314–319.

Insperger, T. and Milton, J.: 2014, Sensory uncertainty and stick balancing at the fingertip, *Biological Cybernetics* **108**(1), 85–101.

Insperger, T., Milton, J. and Stepan, G.: 2015, Semidiscretization for Time-Delayed Neural Balance Control, *SIAM Journal Applied Dynamical Systems* **14**(3), 1258–1277.

Kowalczyk, P., Glendinning, P., Brown, M., Medrano-Cerda, G., Dallali, H. and Shapiro, J.: 2012, Modelling human balance using switched systems with linear feedback control, *Journal of The Royal Society Interface* **9**(67), 234–245.

Lakshmanan, M. and Senthilkumar, D. V.: 2010, *Dynamics of Nonlinear Time-Delay Systems*, Springer-Verlag, Berlin.

Lubashevsky, I.: 2017, *Physics of the Human Mind*, Springer, Cham.

Milton, J. G.: 2013, Intermittent Motor Control: The "drift-and-act" Hypothesis, *in* M. J. Richardson, M. A. Riley and K. Shockley (eds), *Progress in Motor: Control Neural, Computational and Dynamic Approaches*, Springer Science+Business Media, New York, pp. 169–193.

Milton, J., Insperger, T. and Stepan, G.: 2015, Human Balance Control: Dead Zones, Intermittency, and Micro-chaos, *in* T. Ohira and T. Uzawa (eds), *Mathematical Approaches to Biological Systems: Networks, Oscillations, and Collective Motions*, Springer Japan, Tokyo, pp. 1–28.

Milton, J., Townsend, J., King, M. and Ohira, T.: 2009, Balancing with positive feedback: the case for discontinuous control, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **367**(1891), 1181–1193.

Ohira, T.: 2007a, Stochastic Time, *AIP Conference Proceedings* **905**(1), 191–194.

Ohira, T.: 2007b, Stochasticity and non-locality of time, *Physica A: Statistical Mechanics and its Applications* **379**(2), 483–490.

Ohira, T.: 2009, Temporal stochasticity and non-locality, *Journal of Statistical Mechanics: Theory and Experiment* **2009**(01), P01032 (17 pages).

Suzuki, T., Miyazawa, T., Kanemoto, S. and Lubashevsky, I.: 2015, Statistical Properties of Human Response Delay: Analysis of Virtual Stick Balancing Experiments, *Proceedings of the 46th ISCIE International Symposium on Stochastic Systems Theory and Its Applications, Kyoto, Nov. 1-2, 2014*, Institute of Systems, Control and Information Engineers (ISCIE), Kyoto, pp. 236–241.

Zgonnikov, A., Lubashevsky, I., Kanemoto, S., Miyazawa, T. and Suzuki, T.: 2014, To react or not to react? Intrinsic stochasticity of human control in virtual stick balancing, *Journal of The Royal Society Interface* **11**, 20140636.

WEB SITE FOR TEACHING PSYCHOPHYSICAL METHODS WITH MOBILE DEVICES

Kenzo Sakurai[1,2], William H.A. Beaudot[2], and Hiroshi Ono[3]
[1] *Department of Human Science, Tohoku Gakuin University, Sendai, Japan*
[2] *KyberVision Japan LLC, Sendai, Japan*
[3] *Centre for Vision Research, York University, Toronto, Canada*
<sakurai@mail.tohoku-gakuin.ac.jp>

We updated our Web site <http://www.yorku.ca/psycho/> called "Precision and Accuracy with Three Psychophysical Methods". It is an interactive tool designed to teach 3 basic methods (method of limits, method of constant stimuli, method of adjustment), and the conceptual differences between precision (JND) and accuracy (constant error). This online tool has been used in university classes with a large number of students for more than 10 years. Because the previous site was built with the Java technology which is not compatible with contemporary mobile operating systems, we upgraded the scripts written in Java to HTML5. The new site still looks and feels like the previous one, and this modern Web technology allows students off-campus access to the website with their own mobile devices such as smartphones or tablets. [Supported by JSPS Grant-in-Aid for Scientific Research (B) Grant Number 25285202 and (C) Grant Number 17K04498.]

THE RELATIONSHIP BETWEEN THE FILLED DURATION ILLUSION AND THE TIME DILATION ILLUSION

Erika Tomimatsu
Faculty of Design, Kyushu University, 4-9-1 Shiobaru, Minamiku, Fukuoka, Japan
Japan Society for the Promotion of Science

Yoshitaka Nakajima
Faculty of Design, Kyushu University, 4-9-1 Shiobaru, Minamiku, Fukuoka, Japan

Mark A. Elliott
School of Psychology, National University of Ireland, Galway
University Road, Galway, Ireland

Hiroyuki Ito
Faculty of Design, Kyushu University, 4-9-1 Shiobaru, Minamiku, Fukuoka, Japan

Takuya Kishida
Graduate School of Design, Kyushu University
4-9-1 Shiobaru, Minamiku, Fukuoka, Japan
Japan Society for the Promotion of Science

When we compare the subjective length of an empty duration delimited by two short sounds with that of a physically equal filled duration of a continuous sound, the filled duration is perceived as longer than the empty duration. The duration of a moving object can be also perceived to be longer than that of a static object. The former illusion is called the filled duration illusion and the latter is called the time dilation. Our purpose was to investigate the relationship between these two illusions, quantitatively. The method (the method of adjustment) and duration conditions (150–900 ms) were common to both illusions, in order to show the amounts of two illusions in a scale. Test stimuli were pure tones, static randomdots, and dynamic randomdots. The amount of illusion increased with the stimulus duration in the time dilation, whereas it was almost constant in the filled duration illusion.

ESTIMATING VISUAL AVERAGE AND NUMERICAL AVERAGE ACROSS TIME

Hiromi Sato[1,2] and Isamu Motoyoshi[3]
[1] *Faculty of Informatics, Kogakuin University*
1-24-2 Nishishinjuku, Shinjuku, Tokyo, Japan
[2] *JSPS Research Fellow, 5-3-1 Kojimachi, Chiyoda, Tokyo, Japan*
[3] *Department of Life Sciences, The University of Tokyo*
3-8-1 Komaba, Meguro, Tokyo, Japan

Humans can estimate global trend of dynamic events. To understand computational principles underlying the perceptual decision upon the temporal trend, we have been investigating human judgments on temporal average of visual features such as orientation and motion (Sato et al., 2013, 2016). The results showed that observers specifically utilize information ∼200–500 ms before making a decision upon the temporal average. However, we here show that when observers were asked to judge the numerical average of serially presented digits, observers utilized information equally over the stimulus presentation. We found similar tendency for the average number of dots in serially shown textures, but not for the temporal average of facial expression. These results indicate that distinct neural processes are involved in estimating the temporal average of visual representation and of numerical representation.

EFFECTS OF COCHLEAR DELAY ON PERCEPTION OF SIMULTANEITY FOR TWO PURE TONES

Satoshi Okazaki and Makoto Ichikawa
Department of Psychology, Chiba University
1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan
<okazaki@chiba-u.jp, michikawa@chiba-u.jp>

Abstract

Sounds are neurally coded with a frequency-dependent delay at the cochlea, which is the earliest stage of the auditory pathway. Because of this cochlear delay, the simultaneous onsets of the tones with different frequencies would be desynchronized in the early neural representation. The present study examined the effect of the cochlear delay on the perception of simultaneity for two pure tones. We measured the point of subjective simultaneity (PSS) for two pure tones with several frequency separations. We compared the obtained PSSs with the predictions of the cochlear-delay model. We found that the obtained PSSs were closed to the physically synchronous point (the simultaneous onset) and significantly different from the points expected by the cochlear-delay model. The present results suggest that the perceived temporal relationship for sounds reflects the physical synchrony, and that the desynchronization at the cochlea is canceled within the auditory pathway.

This study investigated possible contributions of the cochlear delay upon the relation between perceptual simultaneity and physical synchrony for two pure tones. Cochlea is the earliest stage of the auditory pathway, which codes the physical tones into neural signal. Such neural coding takes place in a tonotopic manner. The high frequency tone is coded near the basal end of cochlea while the low frequency tone is coded near the apex of cochlea. The coding of the tones progresses from the basal end to the apex of cochlea with a time course. This means that the delay in cochlear coding for lower frequency tone is larger than that for the higher frequency tone. Therefore, in accordance with the cochlear delay, the simultaneous onsets of the tones with different frequencies should be desynchronized.

Wojtczak, Beim, Micheyl, and Oxenham (2012) did not find significant effects of cochlear delay on PSS. However, their insignificant results only provide the weak evidence of the effect of cochlear delay. In order to examine whether the cochlear delay affects the perception of simultaneity for two pure tones, this study compared the PSSs with the predictions of the cochlear-delay model based on the statistical significance.

Method

In the experiment, nine listeners participated (three females and six males, 19–32 years of age). The experiment took place in the sound proof room. The stimuli consist of two pure tones with frequency separation (Δf) and onset asynchrony (Δt). Figure 1 shows the schematic spectrogram of the stimuli. The frequency separation conditions were 3, 4, 5, and 6 octave. The higher tone frequency was fixed at 6400 Hz, whereas the lower tone frequency was either 800 Hz, 400 Hz, 200 Hz, or 100 Hz for the respective frequency separations.

Fifteen tone-onset asynchronies (Δt in Fig. 1) were prepared: $\pm$ 0, 2, 4, 7, 14,

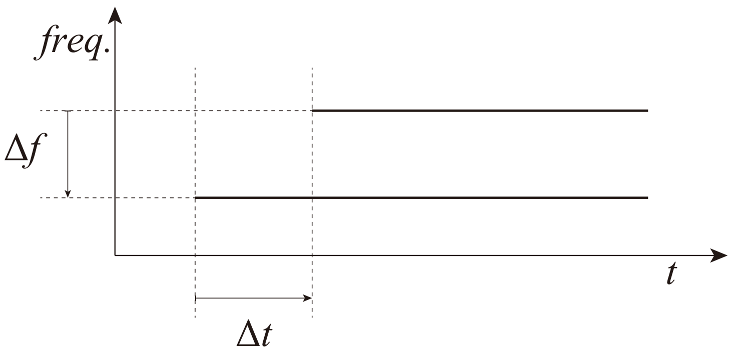


Fig. 1. Schematic spectrogram of stimuli.

27, 52, and 100 ms (positive asynchronies indicate that the tone with lower frequency preceded the tone with higher frequency whereas negative asynchronies indicate that the tone with lower frequency was delayed from the tone with higher frequency). The offset of the tones was always simultaneous. The lagging tone duration was 300 ms, whereas that of the leading tone was lengthened with the absolute value of onset asynchrony (300 ms + onset asynchrony). This 300 ms duration was sufficient to avoid changing the tone loudness because of the variation of tone duration (Florentine, Popper, & Fay, 2011). Tone sound level was set at 35 dBA to avoid the reduction of the amount of cochlear delay due to high sound level (Ruggero & Temchin, 2007). Each tone was tapered with a rise-fall time of 15 ms by the cosine function to avoid spectral splatter at tone onset and offset.

Stimuli were diotically presented via headphones. After the presentation, listeners were asked first to judge whether the two tones were perceptually fused or separated. Only if they judged that the tones were separated were they asked to judge whether the two tones were simultaneous or not. These procedures were intended to avoid listeners'

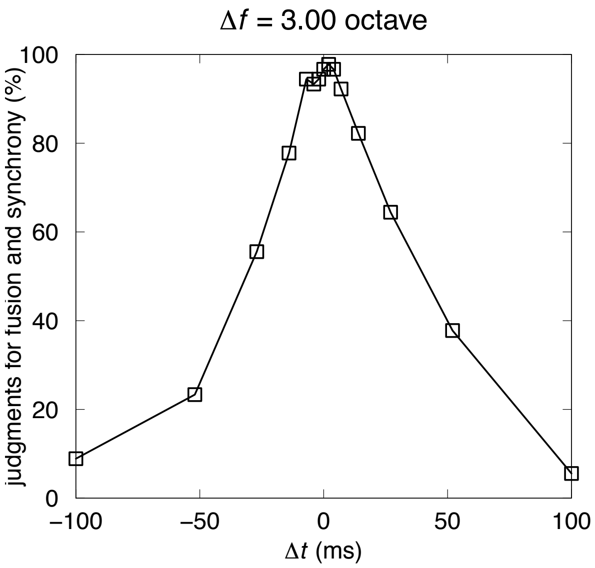


Fig. 2. Distribution of simultaneity judgment for $\Delta f = 3.00$ octave.

judging of the simultaneity based on the perceptual fusion cue (Okazaki & Ichikawa, 2017). Each Δf condition for 15 Δts was presented 10 times in random order. Listeners had no feedback or time pressure for their judgment.

Results and Discussion

Figure 2 shows mean distribution of simultaneity judgment (summed frequency of judgments about fusion and, simultaneity under separate condition). The mean, mode, and median of the distribution were used as indices of PSS for each Δf condition for each listener. The position of mean would represent the center of the probability of simultaneity judgment while that of mode and median would respectively represent the maximum simultaneity and center of the frequency of simultaneity judgment. These indices of PSS were compared to the predictions of the cochlear-delay model (de Boer, 1980; Eggermont,

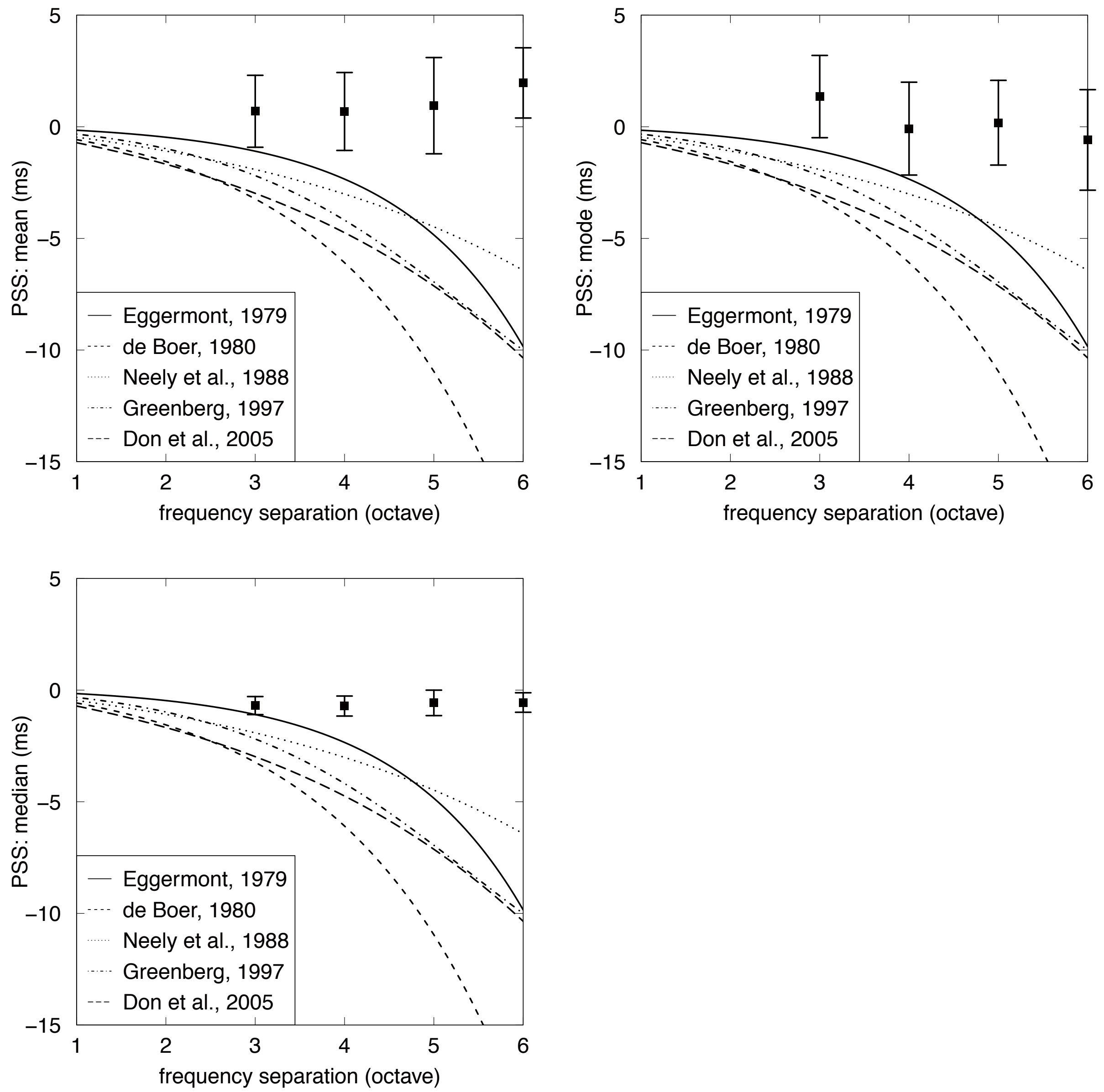


Fig. 3. PSS and prediction by cochlear delay models.

1979; Greenberg, 1997; Neely, Norton, Gorga, & Jesteadt, 1988). The obtained PSS and the predictions were shown in Fig. 3. Error bars show the 95% confidence interval. The PSSs were deviated from the cochlear-delay predictions toward zero asynchrony. Especially, the models predict large effect of cochlear delay at large frequency separation ($\Delta f \geq 5$ octave), and the obtained PSSs were significantly apart from the prediction.

The present results revealed that the cochlear delay has no effect on PSS. The obtained PSSs were closed to the zero asynchrony, physically synchronous point. These results suggest that the cochlear delay should be canceled within the auditory pathway after the cochlea coding. Such cancellation is ecologically valid because if the cochlear delay is reflected to the perception, we would be always wrong for the temporal relationship of sounds with different frequencies.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Numbers JP16J07652, JP25285197.

References

de Boer, E. (1980). Auditory physics. Physical principles in hearing theory. I. *Physics Reports*, 62, 87-174.

Eggermont, J. J. (1979). Compound action potentials: tuning curves and delay times. *Scandinavian Audiology. Supplementum.* 9, 129-139.

Florentine, M., Popper, A. N., & Fay, R. R. (2011). *Loudness.* New York: Springer. p.233.

Greenberg, S. (1997). The significance of the cochlear traveling wave for theories of frequency analysis and pitch. In E. R. Lewis, E. Hecht-Poinar, G. R. Long, R. F. Lyon, P. M. Narins, and C. R. Steele (Eds.), *Diversity in Auditory Mechanics.* Singapore: World Scientific Publishing.

Neely, S. T., Norton, S. J., Gorga, M. P., & Jesteadt, W. (1988). Latency of auditory brain - stem responses and otoacoustic emissions using tone-burst stimuli. *The Journal of the Acoustical Society of America*, 83, 652-656.

Okazaki, S., & Ichikawa, M. (2017). Perceptual simultaneity range as a function of frequency separation for two pure tones. *Acoustical Science and Technology*, 38, 185-192.

Ruggero, M. A., & Temchin, A. N. (2007). Similarity of traveling-wave delays in the hearing organs of humans and other tetrapods. *Journal for the Assocciation Research in Otolaryngoly,* 8, 153-166.

Wojtczak, M., Beim, J. A., Micheyl, C., & Oxenham, A. J. (2012). Perception of across-frequency asynchrony and the role of cochlear delays. *The Journal of the Acoustical Society of America,* 131, 363-377.

THREE-DIMENSIONAL SHAPE RECOGNITION BY ULTRASONIC BINAURAL ECHOES

Miwa Sumiya and Taito Banda
Faculty of Life and Medical Sciences, Doshisha University, Kyotanabe, Japan
Kaoru Ashihara
Human Informatics Research Institute
National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan
Kohta I. Kobayasi
Faculty of Life and Medical Sciences, Doshisha University, Kyotanabe, Japan
Kazuki Yoshino, Masaki Gogami, and Yoshiki Nagatani
Department of Electronic Engineering, Kobe City College of Technology, Kobe, Japan
Yoshiaki Watanabe and Shizuko Hiryu
Faculty of Life and Medical Sciences, Doshisha University, Kyotanabe, Japan

Abstract

Bats use frequency-modulated (FM) and/or constant-frequency (CF) ultrasound for echolocation. In particular, the FM ultrasound is used by the echolocating bats to recognize shape in addition to measure the distance to the object. In this study, we examined if the sighted echolocation novices could discriminate the roundness of edge contours by using the down FM ultrasound mimicking the echolocation sound of the bats. First, our results showed that the all participants could discriminate the differences in the roundness of edge contours among five targets (Targets 1–5 whose corner radii were ranging from 0 cm to 4 cm) by using down FM echoes in the high-frequency range (35–7 kHz) converted the pitch to the audible range (5–1 kHz). In addition, the generalizations were observed in the 3D shape discrimination because the participants could discriminate in all target pairs without training. The participants performed better by using the broadband ultrasounds, i.e., FM sound, harmonic FM sound, and band-limited noise in the high-frequency range than the CF sounds under the conditions which the loudness among the targets were standardized. This suggests that it is important to design the time-frequency structure of the echolocation sound for practical use because it is difficult to use the loudness as the cue in the actual environment.

Bats and dolphins emit ultrasounds and listen to echoes from objects. This is called echolocation. Echolocation in the high-frequency range (ultrasound) is efficient for detecting small objects (their prey) because the sensing resolution depends on the wave length of the echolocation sound. The bats and dolphins use echolocation not to only detect and localize targets, but also to discriminate and identify targets in their respective environment (Simmons et al., 1974; Hammer et al., 1980).

On the other hand, some blind people also use echolocation like bats and dolphins. However, the echolcoating blind people usually use mouth click as echolocation sound because human cannot produce and listen to the ultrasounds. Some previous studies carried out psychophysical examinations which the subjects emit mouth clicks actively to determine the ability in perceiving the location (azimuth), distance, and size of sound-reflecting surfaces (Thaler and Goodale, 2016). For example, the individual who had been blind from birth and had learned to echolocate early in life could detect a change as small as 4° in the azimuthal position of a 150 cm tall pole in a two-interval two-alternative forced choice task (Thaler et al., 2011). However, the long-term training is necessary to master such an echolocation using mouth click. In order to overcome this problem, blind mobility aids have developed based on echolocation system using ultrasounds like bats and dolphins (e.g., Ifukube et al., 1991; Sohl-Dickstein et al., 2015).

We have been conducting bio-sonar research in the echolocating bats. In our previous study, we developed a new system for binaural recording in the ultrasonic range (Uchibori et al., 2015). The system consists of a miniature dummy head (MDH) printed based on the three-dimensional (3D) shape data of a standard dummy head, and a device that converts ultrasonic echoes into audible sounds (Uchibori et al., 2015). It was found that the sounds captured by the proposed MDH system were localized more correctly and outside the head more than normal stereo sounds were. Therefore, we conceived that it might be useful to experience echolocation by MDH.

In this study, we examined if the sighted echolocation novices could discriminate the 3D shape by listening to echoes from the objects. We compared the discrimination performance using stereo echoes in the low- and high-frequency ranges to examine the utility of the high-frequency ultrasound in the discrimination of the roundness of the edge contours. In addition, we discussed effective signal design for perception of the roundness of the edge contours based on the discrimination performance using ultrasonic binaural echoes measured by MDH.

Materials and Methods

Participants

A total of 10 participants P1–10 aged between 20 and 28 (23.0 ± 2.7) took part in the 3D shape discrimination experiment at Doshisha University. Hearing tests revealed that pure tone thresholds of all participants were within 30 dB hearing level up to 8 kHz. All participants reported not to have prior experience with echolocation and they reported to have the normal vision or corrected normal vision. The testing took place in a sound-attenuated chamber (2.3 m (H) × 1.6 m (L) × 1.4 m (W)). Testing procedures were approved by the University ethics board and participants provided with informed consent.

Sound Stimuli

Figure 1A shows custom made five targets, Targets 1–5 were used as sound reflecting objects in this study. These targets were made of solid Narra-wood (Pterocarpus indicus). The surfaces of targets were smooth. As shown in Fig. 1A, it seems that these all five targets, Targets 1–5 were almost same rectangles in a vision seeing from the direction of the Z axis which depends on shadows whereas these five targets have different edge contours. The corner radius increases from 0 to 4 cm with the step of 1 cm for the Targets 1–5 as shown in Fig. 1B. All of the target length were 80 cm. Therefore, the lateral plane areas decrease from 640 to 0 cm2 with the step of 160 cm2 for the Targets 1–5.

We measured the impulse response when the sound was emitted toward the each target by the logarithmic time stretched pulse (Log-TSP) (Stan et al., 2002) in order to compare the transfer function of the targets in the entire frequency band. Measurements were conducted in an anechoic chamber (3.6 m (W) × 4.4 m (L) × 3.4 m (H)) at Doshisha University. The direct sound from the loudspeaker (S-300HR; TEAC CORPO-

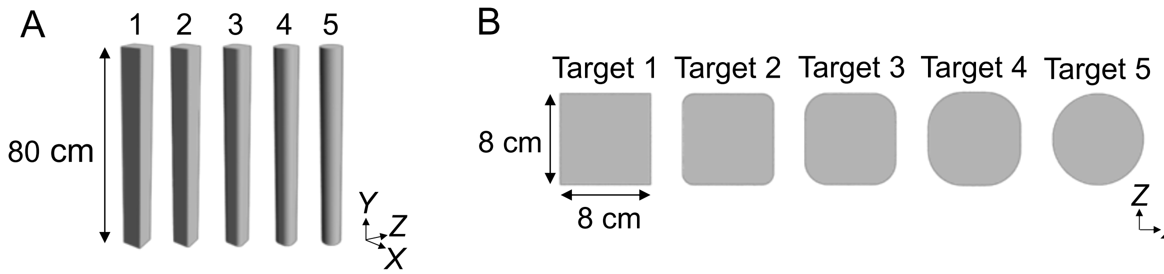


Fig. 1. (A, B) Schematic diagrams of 3D views (A) and top views (B) of the five targets, Target 1–5 having different roundness of the edge contours.

RATION, Tokyo, Japan) and the echoes reflected from each target positioned at 15 cm from the loudspeaker were measured (32-bit accuracy at a sampling rate of 192 kHz) by stereo recording using two omnidirectional condenser microphones of 2.7 mm in diameter (B6 omnidirectional lavalier; Countryman Associates, Inc., California, USA). We also measured the impulse response by monaural recording without the target. The sound pressure level of Log-TSP was 95 dB at 15 cm from the loudspeaker. The background noise level was approximately 14 dB (A) which was low enough to measure the echoes from the targets.

The impulse response was calculated by discrete Fourier transform after averaging signals 32 times by MATLAB (MATLAB; The MathWorks, Inc., Massachusetts, USA). The total frequency characteristics of loudspeaker and microphones were canceled by taking away the impulse response of reference signal measured by the monaural recording from the comparison signal measured by the stereo recording. We used down FM sweeps in low- (5–1 kHz) and high- (35–7 kHz) frequency range for convolution operation with the impulse response of each target. These two signals were convoluted with the echo part which was cut from the entire impulse responses measured by the stereo recordings with the distance between the two microphones of 15.4 cm (5–1 kHz) and 2.2 cm (35–7 kHz), respectively. The high-frequency convoluted signals (35–7 kHz) were 1/7-times pitch converted to audible sounds by time-expansion method using Adobe Audition CC 2015. This allows listeners to listen to the pitch-converted stereo sounds including the acoustic information in the high frequency range as audible sounds. Each sound stimulus consisted of ten successive echoes in a row with the time interval of 35 ms.

Procedure

Randomly-chosen two different sound stimuli were presented in two-alternative forced-choice task by custom-made program of EXPLAB (free software for computational experiment) through headphones (MDR-CD900ST, Sony, Tokyo, Japan). The time interval between the sound stimuli, i.e., inter-stimulus interval was 300 ms. The loudness level of sound stimuli in Target 1 between low- and high-frequency signal conditions were standardized based on the ITU-R BP1770-3 loudness. The sensitivity difference between the left and right microphones was also corrected. All sound stimuli were presented to listeners below 65 dB SPL at the position of eardrum membrane through headphones. The participants were required to answer which sound stimulus was the echoes from the target having more rounded edge contours as soon as possible correctly by pressing the key of the PC.

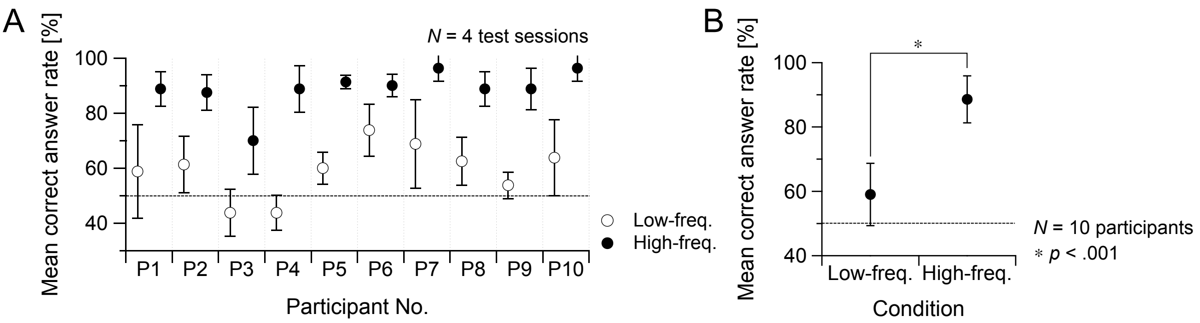


Fig. 2. (A, B) Mean correct answer rate (± SD) of all four test trials in each participant (A) and mean correct answer rate (±SD) of the 10 participants (B) under the low-frequency and high-frequency signal conditions, respectively. The horizontal dashed lines at mean correct answer rate = 50% indicate chance performance. The asterisk highlights significant differences between low-frequency and high-frequency signal conditions.

We conducted 4 test sessions (without feedback of the answer on the display) after 2 or 4 training sessions (with feedback of the answer on the display) per a participant. Here, the participants whose mean correct answer rate was less than 75% in the first 2 training sessions were required to do the additional 2 training sessions. In the training sessions, we used the echoes from only Targets 2 and 4 as the sound stimuli whereas the echoes from all 5 targets, Targets 1–5 were used in the test sessions. There were a total of 28 trials (2 target pairs × 2 signals × 7 repetitions) and 40 trials (20 target pairs × 2 signals × 1 repetitions) in the training and test sessions, respectively.

Results

The mean correct answer rate of the four test sessions in all participants were shown in Fig. 2A. All participants performed better under the high-frequency signal condition (open plots) than under the low-frequency signal condition (filled plots). The paired t-test revealed that mean correct answer rate all ten participants under the high-frequency signal condition was significantly higher than under the low-frequency signal condition (low-frequency, 59.0 ± 9.7%; high-frequency, 88.6 ± 7.3%; the paired t-test, $t(9) = 12.6, p < 0.001$). Furthermore, the mean correct answer rate under the low-frequency signal condition was statistically indistinguishable from the chance level (the paired t-test, $t(9) = 2.9, p = 0.016$).

Frequency spectra of echo parts of the impulse responses under the high-frequency signal condition showed more complex notch pattern than under the low-frequency signal condition (Fig. 3). The differences among the five targets in these frequency spectra were also more remarkable under the high-frequency signal condition (Fig. 3B) than under the low-frequency signal condition (Fig. 3A). These results suggest that the high-frequency signals are more efficient for discrimination of the roundness of edge contours than the low-frequency signals because the complex frequency spectra of the high-frequency echoes which contain the 3D shape information.

Table 1 shows that the participants performed above the chance level in any pair of targets under the both low- and high-frequency signal conditions. This means that the participants could discriminate the roundness of edge contours without training, suggesting generalizations were observed in the 3D shape discrimination by human echolocation.

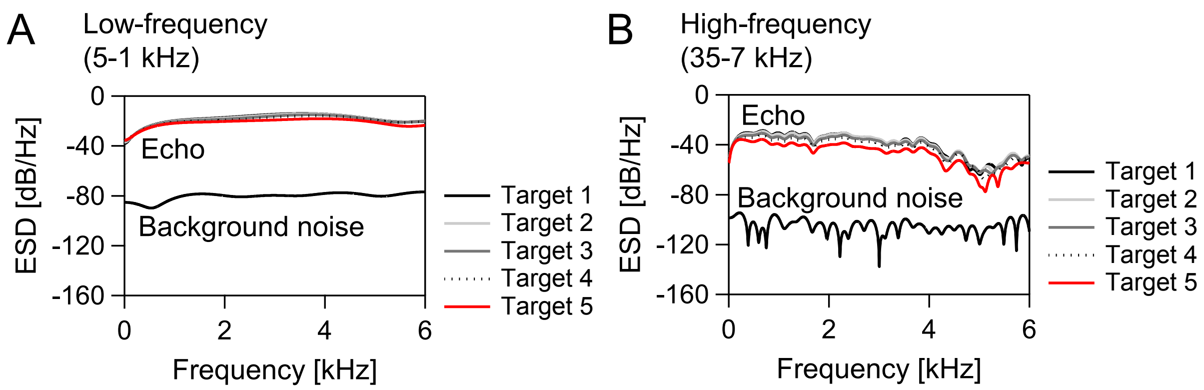


Fig. 3. (A, B) Frequency spectra (energy spectral density) of the echo parts of the impulse response of Target 1–5 (Target 1, black line; Target 2, light gray line; Target 3, dark gray line; Target 4, dashed line; Target 5, red line) which were measured by Log-TSP. The pitch were converted of 1/7 times by time expansion method in B.

Table 1. Mean correct answer rate of the 10 participants (P1–10) in each target pair under low-frequency (left panel) and high-frequency (right panel) signal conditions, respectively.

5-1 kHz, *N* = 10

	Target 1	Target 2	Target 3	Target 4	Target 5
Target 1		50.0	57.5	63.8	71.3
Target 2			63.8	60.0	56.3
Target 3				52.5	58.8
Target 4					61.3
Target 5					

[%]

35-7 kHz, *N* = 10

	Target 1	Target 2	Target 3	Target 4	Target 5
Target 1		68.8	86.3	95.0	93.8
Target 2			81.3	90.0	96.3
Target 3				88.8	97.5
Target 4					88.8
Target 5					

[%]

The mean correct answer rate under high-frequency signal condition was higher than the low-frequency signal condition through all target pairs.

Discussion

In this study, we conducted the 3D shape discrimination experiment using the down FM stereo echoes in the low- and high-frequency ranges. The all participants performed better under the high-frequency condition than under the low-frequency condition (Fig. 2). This indicates that the high-frequency signals are efficient for sensing of the detailed shape information which might be difficult to be discriminated by vision without the shading information. In addition, under the high-frequency signal condition, the participants could discriminate in all target pairs, suggesting that the sighted echolocation novices could generalize the roundness of edge contours by the down FM sound in the high-frequency range.

Because the high-frequency signal was shown to suit for discrimination of the roundness of the edge contours, we compared the discrimination performance using high-frequency signals having different time-frequency structures. We conducted the discrimination tests using eight kinds of binaural echoes measured by 1/7 scaled MDH, i.e., down FM sweep (35–7 kHz), up FM sweep (7–35 kHz), down harmonic FM sweep (35–7 kHz), up harmonic FM (7–35kHz), band-limited noise (7–35 kHz), CF bursts (7 kHz, 21 kHz,

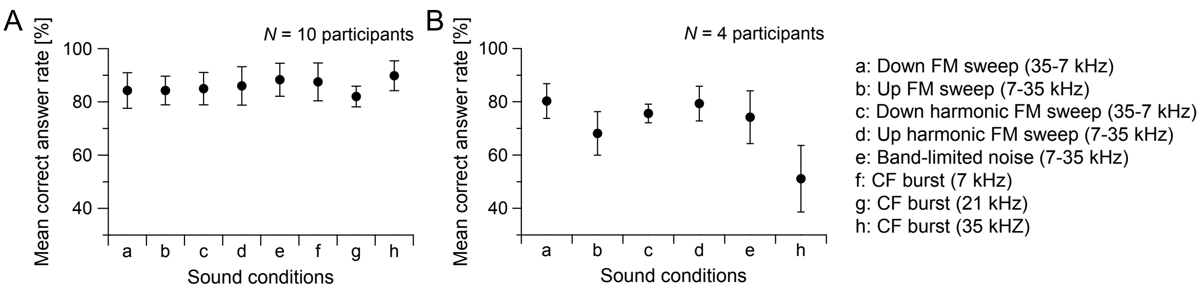


Fig. 4. (A, B) Mean correct answer rate (± SD) of the participants under eight (A, a–h, $N = 10$ participants, P11–20) and six (B, a–e and h, $N = 4$ participants, P21–25) signal conditions in the discrimination experiment using echoes whose loudness were not corrected (A) and using echoes whose loudness were corrected (B), respectively.

35 kHz) for other 10 participants P11-20. These echoes were 1/7-times pitch converted to audible sounds by time-expansion method. According to our previous study, the listeners can perceive 3D sounds when they listen to the pitch-converted binaural sounds through headphones (Uchibori et al., 2015). Figure 4 showed that the mean correct answer rates ranged from 80 to 85% regardless of the signal condition. A generalized linear mixed model (GLMM) with a Binomial distribution and logit link function suggested that the differences between the loudness in each target pair had a significant positive effect on the discriminate performance (a, $\beta = 0.581 \pm 0.098, z = 5.933, p < 0.001$; b, $\beta = 0.620 \pm 0.102, z = 6.104, p < 0.001$; c, $\beta = 0.525 \pm 0.094, z = 5.589, p < 0.001$; d, $\beta = 0.662 \pm 0.113, z = 5.852, p < 0.001$; e, $\beta = 0.737 \pm 0.132, z = 5.572, p < 0.001$; f, $\beta = 0.617 \pm 0.114, z = 5.423, p < 0.001$; g, $\beta = 0.619 \pm 0.095, z = 6.487, p < 0.001$; h, $\beta = 0.489 \pm 0.108, z = 4.527, p < 0.001$). However, these few participants reported that discrimination between the targets was based on the presence of the lower pitch in the target whose edge contours were rounder than the other one.

In order to investigate the performance without the loudness cue, we additionally conducted the discrimination experiments (the loudness level was corrected to the same level) for other 4 participants P21–25 using the echoes from four targets, Targets 1–4 changing the roundness of edge contours (Target 1, cube; Target 4, cylinder) like Targets 1–5 (Fig. 1) used in this study. As a result, the participants performed well above chance under 5 signal conditions (Fig. 4B, a–e) except the CF burst (Fig. 4B, h) condition. GLMM suggested that the differences between the loudness in each target pair did not have a significant effect on the discriminate performance (a–e and h, $\beta = 0.000$) because there were no differences of the loudness among the four targets. The all participants reported to use the pitch information to discriminate. These suggest that the broadband signals in the high-frequency range were useful for the 3D shape discrimination in the environment where we cannot use the loudness cue. Furthermore, the mean correct answer rate under the down FM sound condition might be higher than the up FM sound condition by the masking effect, suggesting that it is also important to select the start frequency in the signal design of the FM sounds. As a matter of course, the discrimination performance might change depending on the relationship between the wave length of the echolocation sound and the size of the corner radius. Therefore, it is also required to design the frequency band depending on the object's size.

In this study, the target was fixed in front of the sound source. However, ultrasonic binaural echoes measured by MDH become more operative in the perception of the moving targets than the fixed targets because the ultrasonic binaural echoes contain the 3D spatial information. Therefore, it is required to extend the discrimination experiment using moving target or targets positioned at the various position in order to apply such an advantage of the ultrasonic binaural echoes.

Acknowledgements

This work was supported by JSPS <https://www.jsps.go.jp> KAKENHI Grant Numbers JP16H06542 (Grant-in-Aid for Scientific Research on Innovative Areas) to SH, and No. 27-3861 (Grant-in-Aid for JSPS Research Fellow) to MS, the Japan Science and Technology Agency (JST) <http://www.jst.go.jp/> PRESTO program to SH.

References

Simmons, J., Lavender, W., Lavender, B., Doroshow, C., Kiefer, S., Livingston, R., Scallet, A., and Crowley, D. (1974). "Target structure and echo spectral discrimination by echolocating bats," *Science*, 186, 1130–1132.

Hammer Jr, C. E., and Au, W. W. (1980). "Porpoise echo-recognition: An analysis of controlling target characteristics," *The Journal of the Acoustical Society of America*, 68, 1285–1293.

Thaler, L., Arnott, S. R., and Goodale, M. A. (2011). "Neural correlates of natural human echolocation in early and late blind echolocation experts," *PloS one*, 6, e20162.

Thaler, L., and Goodale, M. A. (2016). "Echolocation in humans: an overview," Wiley Interdiscip Rev Cogn Sci 7, 382–393.

Ifukube, T., Sasaki, T., and Peng, C. (1991). "A blind mobility aid modeled after echolocation of bats," *IEEE Transactions on Biomedical Engineering*, 38, 461–465.

Sohl-Dickstein, J., Teng, S., Gaub, B. M., Rodgers, C. C., Li, C., DeWeese, M. R., and Harper, N. S. (2015). "A Device for Human Ultrasonic Echolocation," *IEEE Transactions on Biomedical Engineering*, 62, 1526–1534.

Uchibori, S., Sarumaru, Y., Ashihara, K., Ohta, T., and Hiryu, S. (2015). "Experimental evaluation of binaural recording system using a miniature dummy head," *Acoustical Science and Technology*, 36, 42–45.

Stan, G.-B., Embrechts, J.-J., and Archambeau, D. (2002). "Comparison of different impulse response measurement techniques," *Journal of the Audio Engineering Society*, 50, 249–262.

IRRELEVANT SOUND EFFECTS WITH LOCALLY TIME-REVERSED SPEECH: REAL PERFORMANCE DIFFERENCE BETWEEN GERMAN AND JAPANESE NATIVE SPEAKERS?

Kazuo Ueda, Akie Shibata, and Yoshitaka Nakajima
Kyushu University

Katharina Rost, Florian Kattner, and Wolfgang Ellermeier
Technische Universität Darmstadt

<ueda@design.kyushu-u.ac.jp>

A statistically significant performance difference (Wilcoxon, $p < 0.01$) was observed between Japanese ($N = 38$) and German ($N = 81$) native participants in the irrelevant sound experiments [Ellermeier et al. (2017). Fechner Day 2017, Fukuoka, Japan] as well as in our previous investigation [Ellermeier et al. (2015). JASA, 138, 1561-1569]. The high performance level of the Japanese participants might have obscured some of the differences due to irrelevant sound conditions seen in the German listeners. A follow-up experiment with 15 Japanese native participants was undertaken (Fig. 1 shows pooled data with the previous ones), to reveal whether Japanese participants are apt to invent mnemonics during the experiment. A survey was conducted just after the experiment. It revealed that most of Japanese participants used some kinds of mnemonics, and that a half of the participants made puns when they tried to memorize digit sequences. The apparent performance difference between the German and Japanese native participants might be caused by prevailing use of mnemonics in Japanese participants.

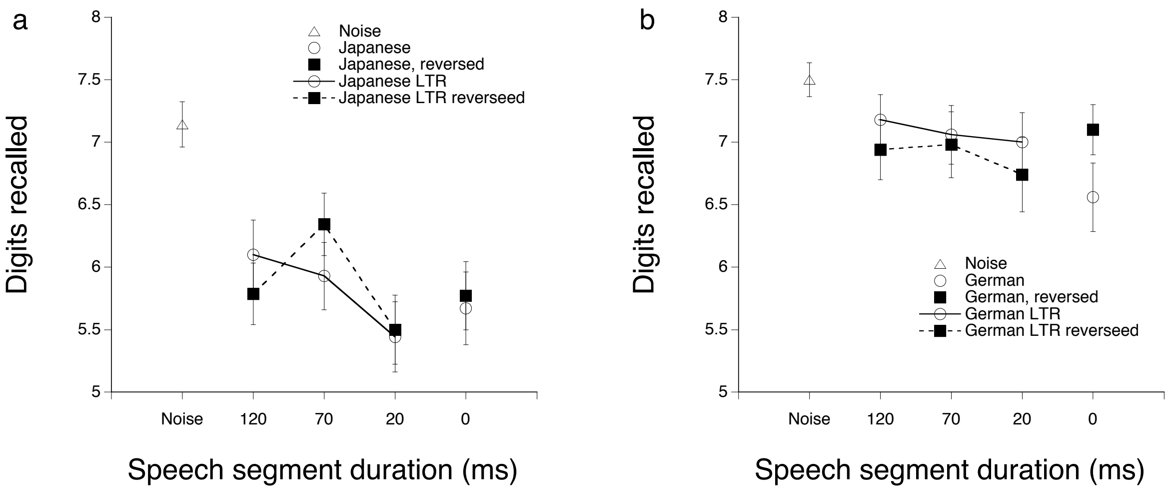


Fig. 1. Performance (average numbers of digits recalled) of native Japanese participants. **(a)** Japanese irrelevant speech was presented ($N = 22$). **(b)** German irrelevant speech was presented ($N = 16$). LTR: locally time-reversed speech. Noise: pink noise. Error bars represent s.e.m.

SPEECH ENCODING SCHEME FOR HEARING AID WITH INFRARED LASER

Yuta Tamai, Kazuyuki Matsumoto, Makoto Arimura, Shizuko Hiryu, and Kohta I. Kobayasi
Neuroethology and Bioengineering Laboratory, Life and Medical Science Doshisha University, 1-3 Tatara Miyakodani, Kyotanabe, Kyoto, Japan
<dmq1041@mail4.doshisha.ac.jp>

Abstract

Action potentials are evoked by irradiating infrared laser to a cochlea. Our group wants to apply the infrared laser to a hearing aid. The purpose of this study was to develop a speech encoding scheme for the hearing aid. The laser stimulation evokes action potentials from entire cochlear nerves simultaneously because it has limited capacity to differentially stimulate cochlear nerves. Thus, the stimulation may create a perception resembling a clicking sound. We created a click modulated speech sound as a simulated sound of the hearing aid. The sound was a repetitive click whose repetition rate and amplitude followed the formant frequency and amplitude envelope transition of an original sound. Subjects listened to the click modulated speech sounds and wrote down how they perceived the sounds. As results, accuracy of vowel was significantly higher than chance. The result suggests that infrared laser stimulation can create at least partially intelligible speech perception.

Cochlear implants are widely used to compensate for hearing loss. The device is hearing aid that provides a sense of sound by stimulating cochlear nerves directly. However, a cochlear implant requires invasive surgery to put on because cochlear implant surgery involves the insertion of electrode into cochlea. The procedure has a possibility of losing remaining hearing.

A previous study revealed that action potentials could be observed by irradiating infrared laser to nerves in vivo (Wells et al. 2005). Because the infrared laser stimulation can stimulate nerves without contacting the tissue, the novel stimulation is paid much attention to as a possible substitute for electrical stimulation. Izzo and her colleagues demonstrated that neural activities were evoked by irradiating infrared laser to cochlear nerves (Izzo et al. 2006)(Izzo et al. 2007)(Izzo et al. 2008). Another study with cats investigated auditory perception created by irradiating infrared laser to cochlear nerves (Matic et al. 2013).

Our group wants to apply infrared laser to hearing aid. The infrared laser can be irradiated from outer canal to cochlear nerves through tympanic membrane. Hence, we can develop noninvasive hearing aid with infrared laser that does not need invasive surgery.

The purpose of this study was to develop stimulating method for creating speech perception by hearing aid with infrared laser. Infrared laser stimulation may create auditory perception like evoked by single-channel cochlear stimulation because laser irradiation evokes almost all cochlear nerves response at the same time. A previous study by Fourcin and colleagues (1979) showed that a single-channel cochlear stimulation was able to produce various acoustic features of speech such as intonation and voiced-voiceless information (Fourcin & Rosen 1979). Other study demonstrated that single-channel cochlear stimulation improved lip-reading ability (Rosen & Ball 1986). However, there are few

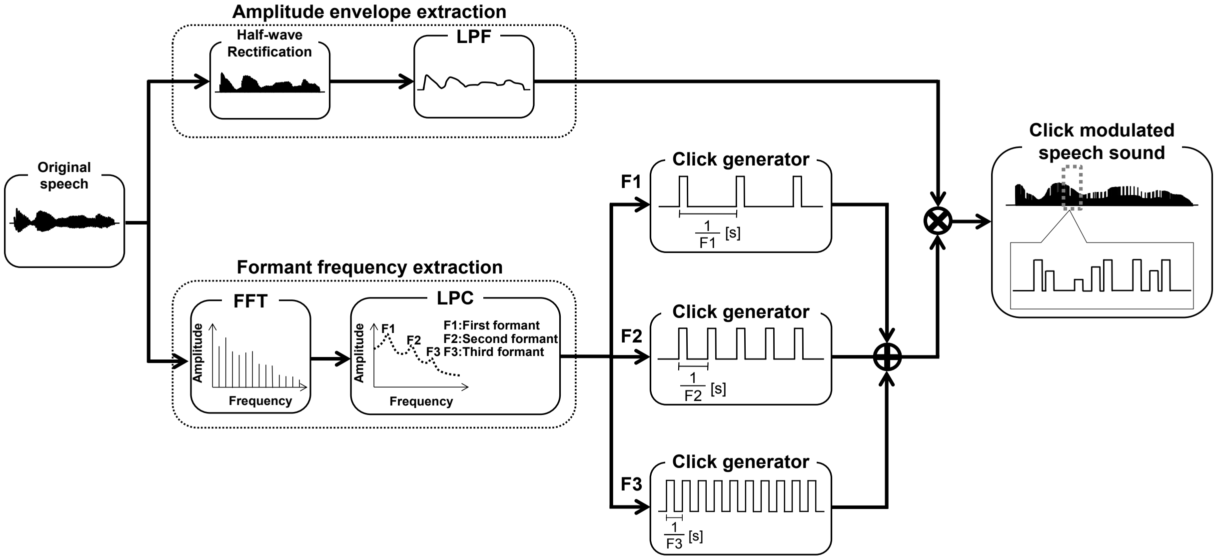


Fig. 1. Encoding scheme of click-modulated speech sound (CMS). The schematic diagram shows how to process the speech signal for synthesizing the CMS. See text for details.

hearing impaired who put on single channel cochlear implant because the intelligibility of single channel cochlear implant was not good enough to produce speech perception clearly.

In this study, we investigated what is the optimal method for creating speech perception with hearing aid with infrared laser. Click-modulated speech sound (CMS) was used as simulated sound of the hearing aid. The sound was a click train, whose pitch (repetition rate) and amplitude followed the formant center frequency and amplitude envelope transition of an original speech sound. We assumed that single-channel cochlear stimulation may create sound perception as like a click sound because a clicking sound stimulated almost all cochlear nerves simultaneously. The CMS is similar to sine-wave speech sound (SWS), which previously Remez and his colleagues developed, because both of sounds replicates formant frequency and amplitude envelope of original sound (Remez et al. 1981). Thus, we assessed that the method for producing intelligible speech perception created by hearing aid with infrared laser by comparing CMS with SWS.

Material and Methods

Subject

Six native Japanese speakers (23–24 years old) participated in the experiment as subjects. All subjects had not listened to the simulated sound of pulsed laser before they became subjects, and passed a hearing screening at 25 HL with frequencies of 0.5, 1, 2, and 4 kHz.

Stimuli: Click-Modulated Speech Sound

We synthesized click-modulated speech sound (CMS). The CMS is a click train, whose repetition rate follows the formant center frequency of an original speech sound, and whose amplitude envelope replicates original envelope. First to Third formant frequency were

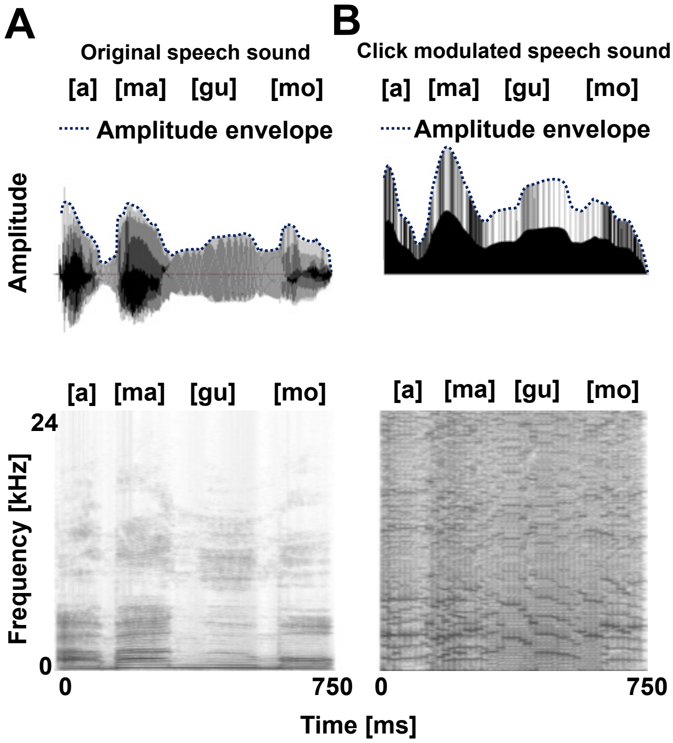


Fig. 2. An example of stimulus. Upper figures show waveform and amplitude envelope (dotted line). Lower figures show spectrum. (A) Original speech sound. A Japanese word "[a][ma][gu][mo]". (B) Click-modulated speech sound synthesized from the original sound (A).

extracted by linear predictive coding (LPC) and fast Fourier transforms (FFTs) at 48 kHz sampling rate and 1024-point FFT length. LPC was calculated every 15 ms over 30 ms Hamming windowed segments. The amplitude envelope was extracted by a low-pass filter (cutoff-frequency = 46 Hz) after half-wave rectification. All signal processing was performed using Matlab (Math Works; Fig.1). An example of an original sound and a click-modulated speech sound was shown in Fig.2. Original speech sounds were Japanese four-mora words voiced by a female speaker. All sounds were from a publicly available data set of familiarity-controlled word lists (FW03) (Amano et al. 2009). We randomly selected 50 words from the various familiarity lists. In this experiment, we used eight kinds of CMS: Constant-frequency (CF), which were average of first formant frequency; first formant frequency (F1); first and second formant frequencies (F1+F2); first, second and third formant frequencies (F1+F2+F3). For all these conditions, the stimulus with amplitude envelope (w/ AE) and without amplitude envelope (w/o AE) were presented.

Stimuli: Sine-Wave Speech Sound

The same words as producing CMS were vocoded to sine-wave speech sound (SWS). The SWS is a sound composed of various sine-waves that follow time-varying formant frequency (Remez et al. 1981). First to third formant frequencies were extracted from original sound in the same method as CMS. We used eight kinds of stimulus as a same as CMS (frequency: CF, F1, F1+F2, F1+F2+F3; Amplitude: w/ AE, w/o AE).

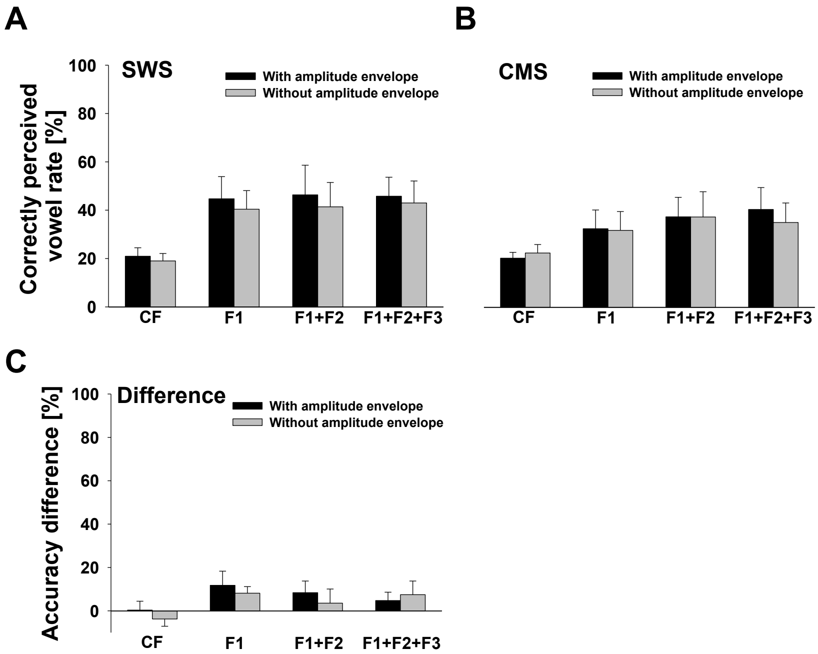


Fig. 3. Intelligibility of click-modulated speech sound and sine-wave speech sound. Error bars indicate standard deviation of the mean. (A) Correctly perceived vowel rate of SWS. (B) Correctly perceived vowel rate of CMS. (C) Accuracy difference between SWS and CMS (SWS-CMS).

Experimental Environment

All experiments were conducted in a sound-proofed room. The stimulus was presented via headphones (STAX Lambda Nove, STAX Industries) through a D/A converter (Octa-capture, Roland). The sound pressure level of all stimuli were measured by microphone (ER-7C Series B, Etymotic Research) and calculated from 60 to 70 dB SPL.

Experimental Procedure

Before an experiment, we presented a speech sound that was not used in the later experiment, and suitable sound pressure level (from 60 to 70 dB SPL) was adjusted for each subject. All subjects participated sixteen sessions. Fifty trials (=50 words) were conducted in each session. The subjects listened to the stimulus and wrote their perception on response sheets using Roman letters within 10 seconds.

Result

Fig.3A shows correctly perceived vowel rates of SWS. For the CF condition, averaged correctly perceived vowel rates were 21% (w/ AE) and 19% (w/o AE). These values were almost the same as a chance level (=20%). For the F1 condition, the values were 45% (w/ AE) and 40% (w/o AE). For the F1+F2 condition, the values were 46% (w/ AE) and 41% (w/o AE). For the F1+F2+F3 condition, the values were 46% (w/ AE) and 43% (w/o AE). For F1, F1+F2 and F1+F2+F3 conditions, the values were higher than chance level. In all conditions, the values were higher when stimuli with amplitude envelope were presented than the values when stimuli without amplitude envelope.

Fig.3B shows correctly perceived vowel rates of CMS. For the CF condition, averaged correctly perceived vowel rates were 21% (w/ AE) and 23% (w/o AE). These values were almost the same as a chance level (=20%) as like the result of SWS. For the F1 condition, the values were 33% (w/ AE) and 32% (w/o AE). For the F1+F2 condition, the values were 38% (w/ AE) and 38% (w/o AE). For the F1+F2+F3 condition, the values were 41% (w/ AE) and 36% (w/o AE). As is the case with the results of SWS, the values were higher than chance level (=20%) for F1, F1+F2 and F1+F2+F3.

Fig.3C shows a difference of correctly perceived vowel rate between SWS and CMS. For CF conditions, a difference was relatively small in w/ AE (0.4%). In w/o AE, the value was -4%. For F1 condition, the values were 12% (w/ AE) and 8% (w/o AE). For F1+F2 condition, the values were 8% (w/ AE) and 4% (w/o AE). For F1+F2+F3 condition, the values were 5% (w/ AE) and 7% (w/o AE).

Discussion

Our results suggested that CMS, which were composed of click train and replicated time-varying formant frequency, were at least partially intelligible as speech sounds. Fig.3A and Fig.3B indicated that speech perception could be created by synthesizing sound composed of either click sounds or sine waves. As many previous studies have demonstrated, formant information is important for speech perception (Peterson & Barney 1952)(Hillenbrand et al. 1995). Remez and his collogues (1981) developed distorted speech sound combing several sine waves, each of which recreated frequency and amplitude of formants (SWS). His group showed that the sound was intelligible (Remez et al. 1981)(Remez et al. 1998). The CMS is similar to SWS in a sense that CMS have information of formant frequency and amplitude envelope of sound. This means that CMS can be perceived as speech sound like SWS. Fig.3C demonstrated that correctly perceived vowel rates of SWS were higher than that of CMS except for CF condition. The results show that SWS was more intelligible than CMS. The difference between SWS and CMS is that CMS has harmonic structure of formant frequencies but SWS does not. Several studies demonstrated that harmonic structures of sound affected auditory perception of both (Moore et al. 1990)(Kewley-Port & Watson 1994). Further study has to pay more attention to the relationship between speech perception and harmonic structure of formant frequencies. Fig.3A and Fig.3B shows that amplitude envelope affects the results of SWS more than that of CMS. Raab and Taub (1969) suggested that intensity discrimination was more difficult with click sounds than longer-lasting tone bursts and noise (Raab & Taub 1969). The phenomenon could relate to the difference between CMS and SWS. Because the result of CMS were rarely affected by amplitude envelope, encoding amplitude envelop into the laser stimulation could have a minor effect on the intelligibility of the stimulation. In this experiment, we quantified the intelligibility of CMS. The sound was developed as simulated sound of hearing aid with infrared laser. When time-varying formant frequencies were replicated, the intelligibility of the sounds was above chance level. In all, our results demonstrated that hearing aid with infrared laser could create speech perception and it could be a new alternative to conventional cochlear implant for restoring speech perception.

Acknowledgements

This work was supported by MEXT KAKENHI Grant Number 15K12069 & 17H01769.

References

Amano, S. et al. (2009). Development of familiarity-controlled word lists 2003 (FW03) to assess spoken-word intelligibility in Japanese. *Speech Communication*, 51, 76-82.

Fourcin, A. & Rosen, S. (1979). External electrical stimulation of the cochlea: clinical, psychophysical, speech-perceptual and histological findings. *British journal of Audiology,* 13(3), 85–107.

Hillenbrand, J. et al. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America, 97(5 Pt 1),* 3099–111.

Izzo, A.D. et al. (2008). Laser Stimulation of Auditory Neurons: Effect of Shorter Pulse Duration and Penetration Depth. *Biophysical Journal,* 94(8), 3159–3166.

Izzo, A.D. et al. (2006). Laser stimulation of the auditory nerve. *Lasers in Surgery and Medicine,* 38(8), 745–753.

Izzo, A.D. et al. (2007). Optical parameter variability in laser nerve stimulation: A study of pulse duration, repetition rate, and wavelength. *IEEE Transactions on Biomedical Engineering,* 54(6), 1108–1114.

Kewley-Port, D. & Watson, C.S. (1994). Formant-frequency discrimination for isolated English vowels. *The Journal of the Acoustical Society of America,* 95(1), 485–496.

Matic, A.I. et al. (2013). Behavioral and Electrophysiological Responses Evoked by Chronic Infrared Neural Stimulation of the Cochlea. *PLoS ONE,* 8(3).

Moore, B.C.J. et al. (1990). Frequency discrimination of complex tones with overlapping and non-overlapping harmonics. The Journal of the Acoustical Society of America, 87(5), 2163–2177. Peterson, G. & Barney, H.L. (1952). Control Methods Used in a Study of the Vowels. *The Journal of the Acoustical Society of America,* 24(2), 175–184.

Raab, D.H. & Taub, H.B. (1969). Click-Intensity Discrimination with and without a Background Masking Noise. *The Journal of the Acoustical Society of America,* 46(4), 965–968.

Remez, R.E. et al. (1998). Multimodal perceptual organization of speech: Evidence from tone analogs of spoken utterances. *Speech Communication,* 26(1), 65–73.

Remez, R.E. et al. (1981). Speech Perception without Traditional Cues. *Science,* 212(22), 947–950.

Rosen, S. & Ball, V. (1986). Speech perception with the Vienna extra-cochlear single-channel implant: a comparison of two approaches to speech coding. *British Journal of Audiology,* 20(1), 61–83.

Wells, J. et al. (2005). Optical stimulation of neural tissue in vivo. *Optics Letters,* 30(5), 504–506.

TEMPORAL RESOLUTION NEEDED FOR AUDITORY COMMUNICATION: MEASUREMENT WITH MOSAIC SPEECH

Yoshitaka Nakajima
Department of Human Science/Research Center for Applied Perceptual Science
Kyushu University, Fukuoka, Japan

Mizuki Matsuda
Nihon Kohden Corporation, Tokyo, Japan

Kazuo Ueda and Gerard B. Remijn
Department of Human Science/Research Center for Applied Perceptual Science
Kyushu University, Fukuoka, Japan

It had been known that speech signals separated into short temporal segments of <100 ms can be remarkably robust in terms of linguistic-content perception against drastic manipulations in each segment, such as partial signal omission or temporal reversal. The temporal resolution needed for speech communication could be far rougher than the well-established temporal acuity of the auditory system: about 2 ms. We developed a new type of degraded speech, mosaic speech, in order to investigate this issue directly. The temporal and the spectral resolution can be degraded independently in this paradigm, and our ultimate aim was to determine the spectro-temporal resolution necessary for speech to become sufficiently intelligible. As a first step, we tried to determine the temporal resolution needed for speech communication.

Two listening experiments ($N = 4$ and $N = 20$) were conducted, and the listeners' intelligibility scores of Japanese speech were obtained with systematically degraded temporal resolutions. This was realized by varying the temporal-window size of mosaic speech from 20 to 320 ms. Locally time-reversed speech, as had been used in previous studies, was also employed for comparison.

Mosaic speech made of short static segments was almost perfectly intelligible with a temporal resolution of 40 ms or finer. Intelligibility dropped for a temporal resolution of 80 ms, but was still around 50%-correct level. The listeners' performance for the mosaic speech was similar to, but significantly better than, their performance for the locally time-reversed speech of the same temporal-window sizes.

Mosaic speech seems a suitable tool to determine the temporal resolution needed for auditory communication—more suitable than locally time-reversed speech. The present data corroborate neuroscientific research indicating that neural oscillations around 30–50 Hz are involved in the segmentation and organization of ongoing speech signals into perceptual units. It was indicated that the human perceptual system can extract meaning from unexpectedly rough temporal information in speech. The process remarkably resembles that of the visual system stringing together static movie frames of ∼40 ms into vivid motion. (This study was supported by the Japan Society for the Promotion of Science and the Center for Clinical and Translational Research, Kyushu University.)

INFLUENCE OF ARTICULATORY SUPPRESSION ON THE CORSI BLOCK TASK: A PSYCHOPHYSICAL ANALYSIS

Shota Mitsuhashi[1]*, Shogo Hirata[2], and Hideyuki Okuzumi[3]
[1]*Graduate School of Education, Tokyo Gakugei University, Japan*
[2]*Department of Elementary Education, Ibaraki Christian University, Japan*
[3]*Faculty of Education, Tokyo Gakugei University, Japan*
*<m121209x@st.u-gakugei.ac.jp>

This study was conducted to investigate the psychological characteristics of the Corsi block task (CBT), a well-known spatial short-term memory assessment, using the dual-task paradigm. The CBT presents a series of nine blocks arranged irregularly. Participants are required to reproduce sequenced tap movements presented by the examiner. In this study, 17 adults (22.3 ± 1.1 years) participated. The computer version of the CBT was applied under three conditions: control, articulatory suppression (AS), and foot tapping (FT). Under the articulatory suppression experimental condition, the phonological loop capacity of the working memory is loaded.

The foot tapping condition shows that the negative effect of articulatory suppression differs from failure of attentional control on a dual task. The presented stimulus sequences (i.e., span) were increased from 2 to 8 in each condition. The participant's correct responses and its time course of movement time (MT) were analyzed. The archived memory spans of each of the three conditions were not different, but psychophysical relations between each serial position and the MT differed only under the AS condition. In the control and FT conditions (Figs. 1 and 3), the MT for the first item of the sequence was significantly longer than the following items. This time course pattern implies the existence of a rehearsal phase before movement execution of the CBT. This tendency, however, was diminished under the AS condition (Fig. 2).

To clarify this observation further, regression analyses were conducted. Each participant's MTs were shown against the serial order. Then, a curve using the reciprocal

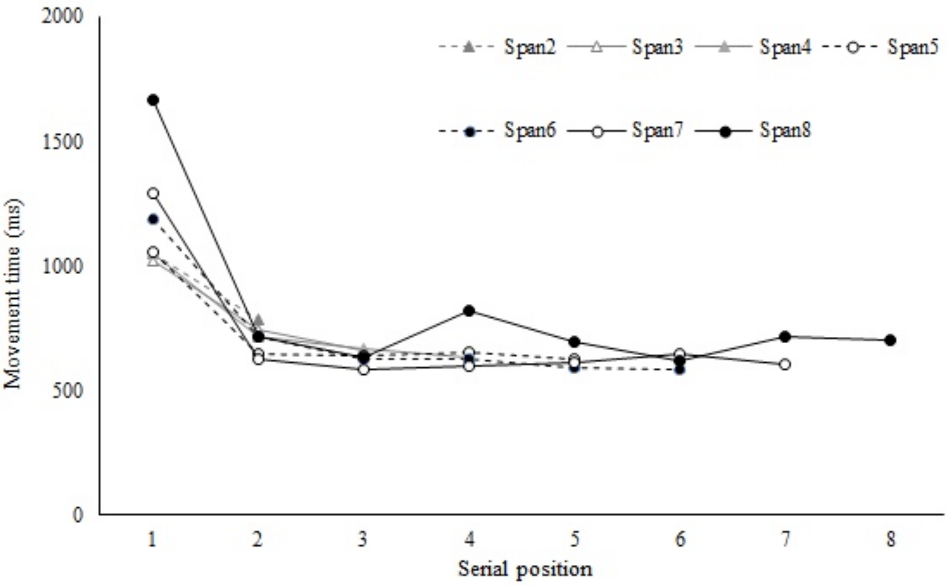


Fig. 1. Each MT under the control condition.

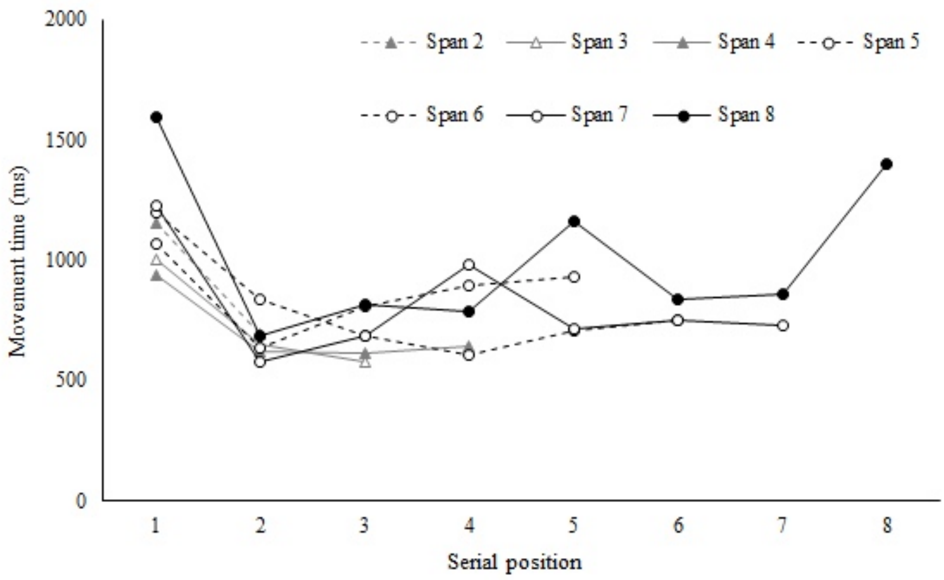


Fig. 2. Each MT under the AS condition.

term was fitted to the MT plot. The coefficient of determination (R^2) was calculated. The respective R^2 of control (0.57 ± 0.21) and FT conditions (0.55 ± 0.23) were significantly higher than that of the AS condition (0.25 ± 0.24; $F_{2,32} = 15.08, p < .01$, partial $\eta^2 = 0.49$). This result implies that articulatory suppression interfered with the existence of a rehearsal phase before movement execution. Therefore, the second rehearsal phase might have occurred at fourth or fifth position of the CBT. The discussion of these results emphasizes the role of the phonological loop of the working memory in facilitating serial recall using visuospatial information.

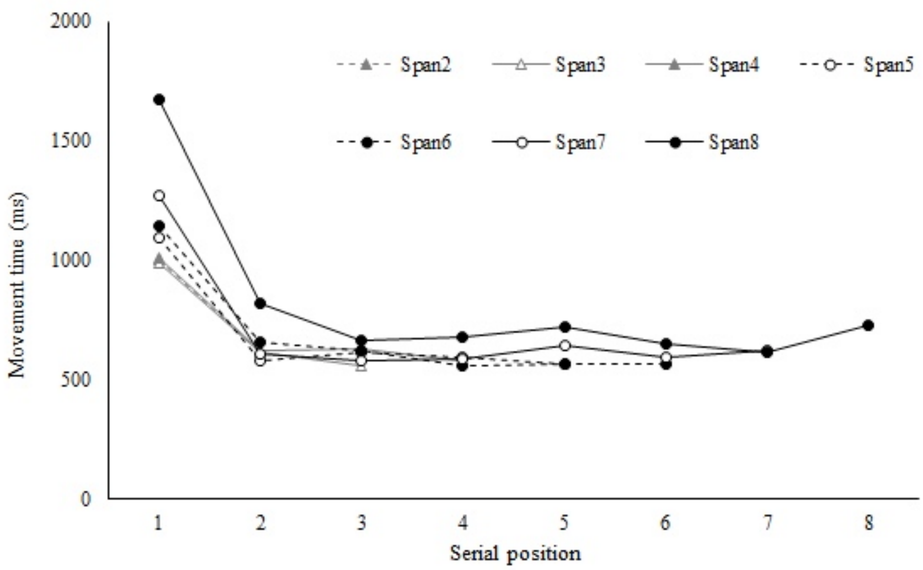


Fig. 3. Each MT under the FT condition.

THE DETECTION TIME FOR HUMAN WITH DIFFERENT ORIENTATION VIEWS IN NATURAL SCENES: EVIDENCE FROM AN ULTRA-RAPID DETECTION TASK

Ryo Tachibana[1] and Keiyu Niikuni[2]
[1]*Department of Psychology, Tohoku University*
Kawauchi 27-1, Aoba-ku, Sendai 980-8576, Japan
[2]*Graduate School of Information Science, Tohoku University*
Aramaki aza Aoba 6-3-09, Aoba-ku, Sendai 980-8579, Japan
<ryo-ta@dc.tohoku.ac.jp, keiyu@cog.is.tohoku.ac.jp>

We investigated how fast and accurately human as an object whose whole body being visible was detected and recognized in natural scenes by manipulating the viewed orientations. Recent studies have shown both human and non-human primate can extract the object information rapidly from a complex situation (i.e., natural scene). In particular, typical animals, such as dog and cat, are recognized with remarkably high accuracy even when presented for quite a short duration (less than 20 ms). However, examining how we process other humans when their whole body is visible in natural scenes has been overlooked, due to the use of only faces for experimental stimuli. Here, we examined how human with different orientation views in natural scenes was recognized, using ultra-rapid detection task that is now a powerful paradigm to examine what is the limit and the time course of object and scene information processing. Examining this issue gives us a new insight into not only the mechanism of objects' perception and recognition but also how it links to social interaction with other people. In Experiment 1, we created two image sets (human in natural scenes as target objects and vehicle in natural scenes as distracter objects) that were based on various image data sets, and evaluated empirically by participants. In Experiment 2, using the image sets as visual stimuli, we examined (1) how fast and accurately human in natural scenes could be detected and recognized, and (2) how human viewed from different orientations in natural scenes affected participants' detection of it using an ultra-rapid human detection task. Results indicated that it took at least 27ms to detect human, and that human was detected more accurately when it was viewed from side than from front and back. This finding suggested that human viewed sideways had more animacy even at low visual levels that could generate human' s shapes and motions biologically when participants watched it in real world.

A COLLECTIVE “PHOTOTAXIS” OF A SOLDIER CRAB

Shusaku Nomura[1] and Haruna Kawai[2]
[1]*Department of Information and Management Systems Engineering*
Faculty of Engineering, Nagaoka University of Technology
1603-1 Kamitomioka, Nagaoka, Japan
[2]*Department of Mechanical Engineering and Robotics, Faculty of Textile*
Shinshu University, 3-15-1 Tokida, Ueda, Japan
<nomura@kjs.nagaokaut.ac.jp, 16bs207a@shinshu-u.ac.jp>

Animal grouping behavior is commonly seen beyond species from tiny insects to mammals. It is always of interest how the animals successfully coordinate their collective behavior with possessing flexibility and plasticity. In this study, an intervention to phototaxis on swarming behavior of a soldier crab, Mictyris guinotae, was experimentally investigated. As for an apparatus, we made “an LED arena” in which LEDs were placed on the wall with 1 cm spaced apart. These LEDs were controlled by microcontroller (Arduino). When a group of crabs was placed into the arena and LEDs were lightened so that the light source rotated along the wall, the crabs followed the light with forming a group. However, when the rate of light rotation got faster than a certain speed, the group of crab rotated in a reverse direction to the light. We discuss the possible mechanism of this intriguing behavior in terms of animal psychophysics.

TEMPORAL CONTINUITY OF VISION AND CYCLIC BINDING OF POSITION AND OBJECT AT THETA RHYTHM

Ryohei Nakayama and Isamu Motoyoshi
Department of Life Sciences, The University of Tokyo
Komaba 3-8-1, Meguro-ku, Tokyo, Japan

Despite the periodical processing of sensory attributes in our brain, the visual system integrates them into conscious continuous perception. To address this issue, we analyzed a novel illusion in which a single Gabor pattern is made to appear temporally discrete. When its spatial window moves constantly whereas the carrier grating stays still or drifts oppositely, the whole stimulus appears to move only intermittently. The apparent rhythm of this discretization is constant at 4–6 Hz regardless of the stimulus speed. Such a slow rate of perceptual updating is similar with temporally fragmented perception reported among patients with damages at MT, and is close to temporal resolution of spatial attention. This rhythmic percept is correlated with the reduction of theta-band neural oscillatory activities around the parietal-occipital cortex. These results suggest that theta-rhythmic periodical neural process related to spatial attention and localization underlies spatial and temporal binding and continuity of visual perception.

DOES EMOTIONAL VALENCE AND AROUSAL AFFECT TEMPORAL RESOLUTION OF VISION?

Misa Kobayashi[1] and Makoto Ichikawa[2]
[1] *Graduate school of Humanities and Public Affairs, Chiba University*
[2] *Faculty of Humanities, Chiba University, Japan*
<ms.kobayashi@chiba-u.jp>

Abstract

Our previous study showed that negative and high-arousal pictures would increase the temporal resolution of the visual processing as well as elongate the subjective duration. In the present study, in order to understand how emotional arousal and valence affect the temporal resolution, we used pictures with wider range of arousal (high, medium and low) and valence (negative and neutral) than previous study did. As an index of temporal resolution of the visual processing, we measured the duration with which observers may detect the monochrome picture in viewing chromatic pictures. We found a significant main effect of valence level, and interaction of arousal and valence levels; the minimum duration for the monochrome picture detection was especially reduced in high-arousal with negative valence while it was extended in high-arousal with neutral valence. Results indicate that arousal is effective in facilitating the temporal resolution of the visual processing when it interacts with valence.

Several studies have shown that emotion may affect the subjective duration in viewing pictures (e.g., Angrilli, Cherubini, Pavese, & Manfredini, 1997; Yamada & Kawabe, 2011). However, those previous studies have not understood whether the effect occurs in the early stage of perceiving the stimuli, or in the latter stage of recalling the memory for the pictures.

In our previous study (Kobayashi & Ichikawa, 2016), we investigated if the emotion affects the temporal resolution of the visual processing as well as subjective duration in viewing the emotional pictures. In that study, we used the pictures, which would evoke either safe or dangerous impressions, from International Affective Picture System (IAPS; Lang, Öhman, & Vaitl, 1988). As an index of temporal resolution of the visual processing, we measured the noticeable duration of a brief monochrome picture in viewing chromatic picture by the use of methods of constant stimuli. Results of the study showed that negative emotion elongated subjective duration, and also improved temporal resolution of the visual processing. Regression analysis revealed that the temporal resolution of the visual processing increased only with the arousal level of the pictures while the subjective duration increased not only with the arousal level, but also with the valence and dominance levels. These results suggested that arousal might be the determinant factor to affect the temporal resolution of the visual processing. However, that study used pictures with the limited range of arousal; we used only high arousal pictures. To understand the effect of arousal upon the temporal resolution of the visual processing properly, we should use pictures with wider range of arousal than the previous study did. In this study, in order to investigate how emotion affects the temporal resolution of the visual processing, we prepared pictures with wider range of arousal (high, medium and low) and valence (negative and neutral) levels than previous studies did. By using these new pictures, we measured the temporal resolution of the visual processing by the use of the same method as in the previous study.

Method

Nineteen observers, eleven females and eight males, took part in this experiment.

Stimuli and Apparatus

A personal computer (Dell Vostro 420) presented stimuli on a 17-inch display (Eizo T561, 100 Hz). The viewing distance was about 57 cm. Observers sat on a chair in front of a desk (80 cm in height), with their head fixed on a chin rest. Stimulus presentation and record of observer responses were conducted by Super Lab ver. 4.5 (Cedrus).

We prepared six conditions of the stimuli by combining of three arousal levels (high, medium and low) and two valence levels (negative and neutral). We chose 10 chromatic pictures for each condition from IAPS. (Means of arousal score in IAPS for high arousal with negative valence, high arousal with neutral valence, medium arousal with negative valence, medium arousal with neutral valence, low arousal with negative valence, and low arousal with neutral valence are respectively 6.64 (0.35), 6.50 (0.36), 4.99 (0.32), 4.95 (0.26), 3.82 (0.18), and 2.88 (0.27) (numerals in parenthesis show SD). Means of valence score in IAPS for each condition are also respectively 2.45 (0.58), 5.07 (0.93), 3.26 (0.55), 4.98 (0.62), 3.38 (0.50), and 5.38(0.52) (numerals in parenthesis show SD)).

The stimulus size was fixed at 9.0 × 10.2 deg. Background of the display was black (.005 cd/m^2, 11.4 × 16.2 deg). We created monochrome pictures from the 60 original pictures with Photoshop ver. 11.0.2 (Adobe).

Procedure

At the beginning of each trial, an instruction to start the trial was presented in white characters (119.329 cd/m^2) at the center of a black background (Fig. 1). When the observer pressed the space key, a white fixation cross (.7 × .7 deg) was presented at the center of the display for 500 ms. Then, one of the original chromatic pictures was presented for 1000 ms. On the two third of the trials, the chromatic picture was followed by a monochrome version of the same picture. There were five duration conditions for the monochrome picture, ranging from 10 to 50 ms by 10 ms steps. On the other one third of the trials, the original chromatic pictures were presented again for the same duration conditions of the monochrome picture. Immediately after the monochrome or chromatic

Fig. 1. A trial format of experiment.

Table 1. The average ratings for each condition (numerals in parenthesis show SDs).

Arousal	High		Medium		Low	
Valence	Negative	Neutral	Negative	Neutral	Negative	Neutral
Arousal	2.04 (.58)	1.42 (.73)	0.29 (.76)	0.06 (.86)	-0.82 (.79)	-1.72 (.54)
Valence	-1.98 (.66)	-0.24 (.71)	-1.20 (.48)	0.04 (.57)	-0.62 (.61)	0.71 (.55)
Safe-Danger	-2.25 (.93)	-1.48 (.53)	-0.57 (.74)	0.06 (.92)	0.30 (.66)	1.66 (.65)

picture presentation, the display turned to a random dot field (9.0 × 10.2 deg), and observers reported whether they saw the monochrome picture by key press.

Each of the 60 pictures was presented with five duration conditions twice for monochrome pictures (600 trials) and once for chromatic pictures (300 trials). Therefore, each observer completed 900 trials. The stimuli were presented in random order.

In order to assess participants' emotion during observing pictures, we asked participants to rate impressions of all pictures by the use of the 7-point bipolar scale (-3.0 to $+3.0$, with 0 as neutral) for pleasantness (pleasant-unpleasant), safety (safe-danger), and impact (shocking- ordinary) after the task of temporal resolution of the visual processing. Each of the 60 pictures was presented once at the center of the black display for 1.0 s in a random order for each observer.

Results

The data of one participant was excluded in the following data analyses because he rated the same arousal level for all of the pictures. The data of two participants were also excluded because they could see the monochrome picture even with 10 ms presentation above chance level.

Table 1 shows the average ratings for each condition. Means of arousal and valence were consistent with what is expected. We performed a Probit analysis (Finney, 1971) to obtain 50% thresholds for the detection of the monochrome versions of each condition for each participant. The χ^2 test found that the data did not fit to a psychometric function for four observers. Figure 2 shows the average of the 50% thresholds of 12 observers, whose data fit to psychometric function for all of the six conditions. A two-way within subject analysis of variance (ANOVA) on the 50% threshold with arousal (high, medium and low) and valence (negative and neutral) as factors was performed. There was a significant main effect of valence, $F(1, 11) = 13.69, p < .01$, while there was not a main effect of arousal, $F(2, 22) = 3.00, p > .05$. The threshold in negative condition was shorter than that in neutral condition. The interaction of arousal and valence was also significant, $F(2, 22) = 4.79, p < .02$. Bonferroni correction was applied for post hoc pairwise significance tests. In the high arousal condition, the minimum duration in which observers could notice monochrome image in viewing negative pictures was shorter than that in viewing neutral pictures ($p < .01$). This result is consistent with the results of our previous study (Kobayashi & Ichikawa, 2016). Also, in the neutral valence conditions, observers needed longer duration to detect monochrome image in viewing high arousal pictures than in viewing medium arousal pictures ($p < .02$). These results suggest that temporal resolution of the visual processing is affected by combination of emotional valence and arousal levels in viewing pictures. The similar result was confirmed with a

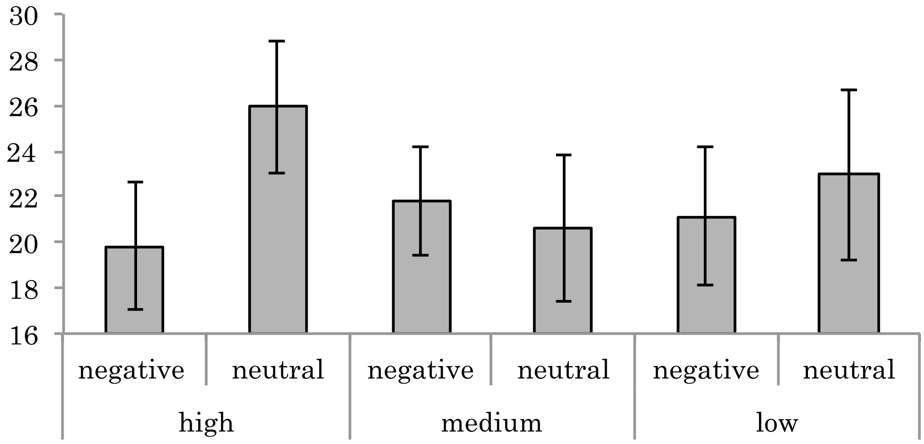


Fig. 2. The average of 50% threshold of each participant for each condition.

stepwise multiple regression analysis [$R^2 = .035, F(3, 956) = 11.427, p < .01$]. All valence, arousal and dominance scores of IAPS significantly correlated with the frequency of a correct report of a monochrome picture ($\beta = -.595, .912, .439$, respectively).

Discussions

In this study, we provided the sufficient range of arousal levels in order to investigate the effect of arousal. The minimum duration in which observers could notice monochrome image was especially reduced in combining high arousal and negative valence. Also, the minimum duration for the monochrome picture detection was most extended in combining high arousal and neutral valence. These results showed that emotion affect the temporal resolution of the visual processing, in a consistent way with that of our previous study (Kobayashi & Ichikawa, 2016).

The present results showed that arousal did not affect the temporal resolution of the visual processing by itself although it interacts with valence. These results indicate that arousal is effective in facilitating the temporal resolution of the visual processing when it interacts with valence. Since this study examined only the effect of emotion upon temporal resolution of the visual processing, the relationship between temporal resolution and subjective duration is still unknown. There is a study about subjective duration suggesting that subjective duration is affected by interaction of arousal and valence (Angrilli et al., 1997), so both the temporal resolution and subjective duration are affected by the combination of emotional arousal and valence. Investigating how emotion, especially combination of arousal and valence, may affect the subjective duration would be helpful to understand the basis for the effects of emotion upon temporal resolution of perception.

References

Angrilli, A., Cherubini, P., Pavese, A., & Manfredini, S. (1997). The influence of affective factors on time perception. *Attention, Perception, & Psychophysics,* 59(6), 972–982.

Finney, D. J. (1971). *Probit analysis.* Cambridge: Cambridge University Press.

Kobayashi, M., & Ichikawa, M. (2016). Emotions evoked by viewing pictures may affect temporal aspects of visual processing. *Japanese Psychological Research,* 58(3), 273–283.

Lang, P. J., Ohman, A., & Vaitl, D. (1988). *The international affective picture system* [photographic slides]. Gainesville, FL: Center for Research in Psychophysiology, University of Florida.

Yamada, Y., & Kawabe, T. (2011). Emotion colors time perception unconsciously. *Consciousness and Cognition,* 20(4), 1835–1841.

COLOR-TEMPERATURE ASSOCIATION IN DICHROMATIC AND TRICHROMATIC INDIVIDUALS

Yuki Mori[1] and Chihiro Hiramatsu[2]
[1]*Department of Visual Communication Design, School of Design, Kyushu University, 815-8540 Fukuoka, Japan*
[2]*Department of Human Science, Faculty of Design, Kyushu University, 815-8540 Fukuoka, Japan*

Associating color with temperature (e.g., red is hot/warm, and blue is cold) is one of the broadly accepted forms of multisensory processing in humans. Previous studies have suggested that such association is based on a loosely held cultural norm. Moreover, individuals learn these conventional associations through experience over the course of development. This evidence raises the question of how individuals with color perceptions different from common trichromatic vision develop color-temperature associations in a certain culture. This study aimed to examine how individuals with dichromatic vision, constituting ∼2% of the male population, associate colors with temperature and to compare the pattern of association to that in trichromatic individuals through psychophysical experiments. We will describe our latest results and discuss the influence of cultural norm on the multisensory processing of individuals who perceive color in modes different from common trichromatic vision.

MEASUREMENT OF VECTION STRENGTH INDUCED BY VECTION SCENES IN THE JAPANESE ANIMATIONS

Kousuke Tokunaga[1], Yoshitaka Fujii[2], Masaki Ogawa[1], Satoshi Ikehata[3], Tomohiro Masuda[4], and Takeharu Seno[1,5]

[1]*Faculty of Design, Kyushu University, 4-9-1 Shiobaru, Minami-ku, Fukuoka, Japan*

[2]*Ritsumeikan University, College of Comprehensive Psychology 2-150 Iwakura-cho, Ibaraki-shi, Osaka, Japan*

[3]*National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, Japan*

[4]*Faculty of Human Sciences, Bunkyo University, 3337 Minami-Ogishima Koshigaya-shi, Saitama, Japan*

[5]*Research Center for Applied Perceptual Science, Kyushu University 4-9-1 Shiobaru, Minami-ku, Fukuoka, Japan*

The visually induced self-motion perception is known as “vection”. In this study, our purpose was to confirm the factors of vection in the Japanese animations. Prior to measurement of magnitude of vection, we collected 150 vection scenes in the Japanese animation into a database. We defined “vection scene” as a scene in which there was a coherent wide range motion that could induce vection. Then, we conducted a psychological experiment to measure vection strength obtained by those scenes. Participants were asked to rate the relative strength of each vection scene as compared to that of the standard pattern (an expanding circular grating motion). In addition, we also analyzed inter-frame variation in each vection scene to reveal the relationship between amounts of temporal characteristic in each scene and the obtained vection strength. Our results suggested various important things in vection in the Japanese animations.

EFFECT OF FRAME RATE ON VECTION STRENGTH

Yoshitaka Fujii[1], Takeharu Seno[1] and Robert S. Allison[2,3]
[1]*Faculty of Design, Kyushu University, Fukuoka, Japan*
[2]*Department of Electrical Engineering and Computer Science*
York University, Toronto, Canada
[3]*Centre for Vision Research, York University, Toronto, Canada*

We examined effect of frame rate of optical flow on vection strength. Downward (Experiment 1) and expanding (Experiment 2) grating movies were used as stimuli to induce vection. Frame rates were controlled in seven conditions (3, 4, 6, 12, 20, 30 and 60), and vection strength were measured with three indices (latency of vection onset, total duration time and subjective magnitude). We hypothesized that higher frame rate should induce stronger vection because low frame rate cause artifacts such as judder and motion blur. The results of both experiment clearly showed that vection strength increased with increasing frame rate, however, the rate of increase were not constant and saturated in the high range.

DEVELOPMENT OF ASYMMETRY IN VECTION FOR EXPANSION/CONTRACTION RADIAL OPTIC FLOWS

Nobu Shirai[1], Shuich Endo[2], Shigehito Tanahashi[3], Takeharu Seno[4,5], and Tomoko Imura[6]

[1]*Department of Psychology, Faculty of Humanities, Niigata University*
[2]*Department of Electrical and Information Engineering Graduate School of Science and Technology, Niigata University*
[3]*Department of Biocybernetics, Faculty of Engineering, Niigata University*
[4]*Faculty of Design, Kyushu University*
[5]*Research Center for Applied Perceptual Science, Kyushu University*
[6]*Department of Information Systems, Faculty of Information Culture Niigata University of International and Information Studies*
<shirai@human.niigata-u.ac.jp>

Vection, illusory self-motion elicited by visual stimuli, is easier to be induced by radial contraction than expansion flows (e.g., Bubka et al., 2008; Seno et al., 2010). The asymmetric feature in vection was reexamined with 17 younger children (M = 7.9 years), 19 older children (M = 10.5 years), and 20 adults (M = 21.4 years). In each experimental trial, participants observed either a radial expansion or contraction flow, and latency, cumulative duration, and magnitude of vection were measured. Results indicated the latency for contraction was significantly shorter than that for expansion in all age groups. Additionally, the latency and the magnitude were significantly shorter and greater, respectively, in the younger and older children than the adults in both expansion/contraction flows. The asymmetry in vection for expansion/contraction emerges at the latest in school age, while the overall sensitivity to vection in school age children is significantly different from that in adults.

POSTURAL INSTABILITY PREDICTS THE LIKELIHOOD, BUT NOT THE SEVERITY, OF CYBERSICKNESS IN HMDs

Stephen Palmisano, Benjamin Arcioni, Juno Kim, Deborah Apthorp, and Paul J. Stapley
School of Psychology, University of Wollongong, NSW, 2522, Australia
<stephenp@uow.edu.au>

In this study we examined motion sickness when users viewed optic flow presented via an Oculus Rift HMD as they continuously rotated their heads left and right. Head tracking was used to shift the user's viewpoint either in an ecological direction or in the completely opposite direction. We obtained pre-test measures postural instability, had participants complete pre-post items on the simulator sickness questionnaire (the SSQ), and asked them whether they felt sick after each experimental trial. Those who responded that they felt sick after trials displayed significantly larger sway areas and greater anterior-posterior sway variability prior to testing (compared to users who always responded that they were well throughout the experiment). While total SSQ scores were significantly higher for "sick" participants, sway measures did not significantly predict symptom type or severity.

VIEWING HAND MOTION CAN ELICIT TICKLE SENSATION

Godai Saito, Reo Takahashi, and Jiro Gyoba
Department of Psychology, Graduate School of Art and Letters, Tohoku University
27-1 Kawauchi, Aoba-ku, Sendai 980-8576, Japan

We tested whether viewing hand motion induce tickle sensation without touching the participant's knee. The participants were asked to rate the degree of their subject feeling of tickle sensation from 0 (not at all tickly) to 10 (extremely tickly) after they watched the titillative hand motion with or without touching their own knee. The results showed that visual hand motion presented near to the knee (within approximately 10 cm from the knee) induced tickle sensation, whereas the hand motion presented far to the knee could not. The tickliness rating increased significantly as the hand approached the participant's knee regardless of the fixation point. Our results suggest that the interactions between visual hand motion and tickle sensation occur in a limited space surrounding body parts.

VISUAL WORKING MEMORY REPRESENTATION OF COMPOUND STIMULI: EFFECTS OF AUDIOVISUAL INTERACTION AT DIFFERENT VIRTUAL DISTANCES

Qiongyu Wang and Lihan Chen*
School of Psychological and Cognitive Sciences, Peking University
Beijing 100871, China
and Beijing Key Laboratory of Behavior and Mental Health, Peking University
*Correspondence: <CLH@pku.edu.cn>

Abstract

In daily life, we perceive multisensory stimuli at different distances, with hierarchical representations of visual working memory at both global outline and local details. However, the precisions of those representations have not been rigorously examined. Here in a virtual reality environment, we presented four Navon letters surrounding a central fixation, either at near (ca. 0.7 m) or far (70 m) distances with same visual angle. Within 1.2 seconds, one element letter (small letter) in each Navon compound (large letter) changed color constantly. And in most trails, one letter changed color with synchronous sound beeps. Participants reported whether the subsequent probe Navon compound was identical to the 'encoded' letter at the same position or not. The result showed an overall global precedence effect, while precision of visual working memory encoding was hampered when the visual object was presented at far distance with asynchronous sounds. This study has ecological implications for studying memory representation in real multisensory scenario.

Visual working memory (VWM), as an interface between perception and long-term memory, allows people to process visual information in real time and form a relatively stable representation of the vision world (Hollingworth, 2004). Recent evidence has shown that human and other primates' behaviors (including memory) benefit from audiovisual interaction (Bigelow & Poremba, 2016). Indeed, multisensory integration during working memory (WM) has attracted little attention, though a few studies have explored the performance of memory retrieval under semantic correspondence between audiovisual events during the encoding stage (Xie, Xu, Bian, & Li, 2017). Therefore, it is of great realistic significance to explore how visual information is represented in VWM under multisensory context, and whether the capacity of VWM goes beyond 4 as assumed (see a review by Luck & Vogel, 2013).

Encoding of visual information is mediated by the hierarchical structure in the visual objects (Nie, Müller, & Conci, 2017), and subserved by the attentional processing (Schmidt, Vogel, Woodman, & Luck, 2002; Jiang, Olson, & Chun, 2000). Hierarchical stimuli, such as Navon letters, provide excellent materials to study how attention is distributed across local-level elements and global object. The global precedence effect has been shown in a number of studies (Navon, 1977; Nie et al., 2017). According to Construal Level Theory (CLT; Trope & Liberman, 2010), people represent objects at different psychological distances with different construal levels. Specifically, if an object is psychologically distant, people will pay more attention to the global feature of the object. On the contrary, people will focus on the local properties of the individual elements (Liberman & Förster, 2009). In multisensory context, varying the distance in depth between different sensory stimuli will change the functional priorities of individual sensory modalities during multisensory interaction, and hence affect the observers' perceptual performance upon the targets (Van der Stoep, Nijboer, Van der Stigchel, & Spence, 2015; Welch & Warren, 1980).

In current study, we presented Navon letters either at near or far distances in a virtual reality environment, with or without concurrent sounds. We hypothesized that auditory influence upon the representation of VWM would differ with respective to the perceived distance in depth between audiovisual stimuli, where the short distance make the observers focus on the fine details of the individual elements while the far distance guide them to pay attention on the global outline. The results showed an overall global precedence effect, while precision of visual working memory encoding was hampered when the visual object was presented at far distance with asynchronous sounds.

Method

Participants

Thirty participants (8 males, mean age = 21.6, range 18–24) took part in the experiment. All reported normal or corrected-to-normal vision and normal hearing. Participants were provided written consent to the procedure of the experiment and were paid ¥40 an hour.

Stimuli and Apparatus

Memory stimuli were four Navon letters, appearing on the top, bottom, left and right to a central fixation. Each Navon letter was a global letter (G, H, O, R, or S, visual angle: 4.58° × 7.85°) made up of several (16–18) identically-shaped, red and green local

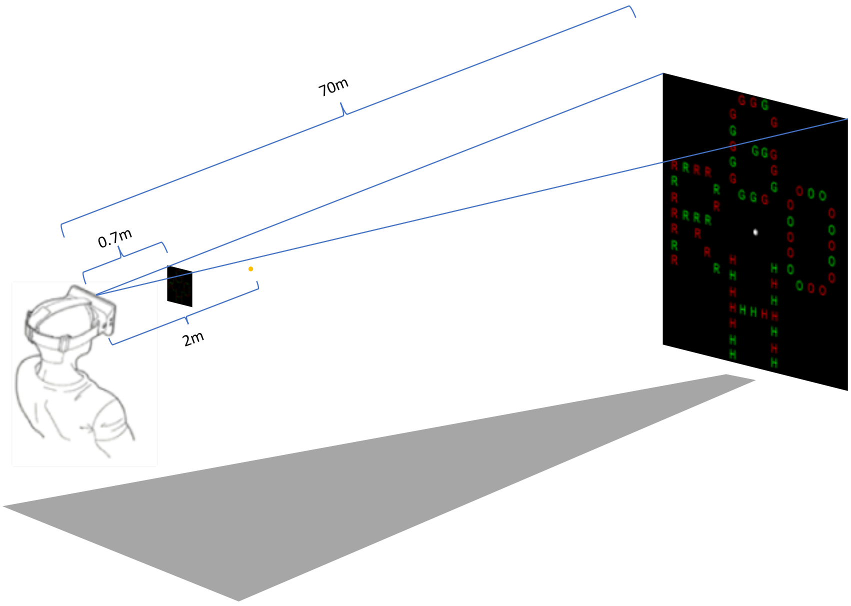


Fig. 1. Illustration of the experiment. Participant was wearing an Oculus Rift headset and memory array was presented at near (0.7 m) or far (70 m) distance with same visual angle. The yellow fixation was presented at 2 m. And there was always a gray distance cue on the bottom of the screen.

letters (G, H, O, R, or S, visual angle: 0.65° × 0.81°), arranged in an invisible 5 × 7 grid. For each trial, four Navon letters were different from each other at both global and local levels. Target (probe) locations were randomly selected in each trial. And if the probe letter changed in global or local level, it changed into one of G, H, O, R, or S which didn't presented in the memory array.

Sound beep was a 500-Hz tone (44.1 kHz sample rate, 16 bit) with a duration of 40 ms (including a 5 ms fade-in and a 5 ms fade-out to avoid clicks).

The virtual environment was implemented using Unity 5.5.1(Unity Technologies). Visual stimuli were presented on a head-mounted display (Oculus Rift, featured a 110° field of view, two 1080 × 1200 OLED screens, with a combined resolution of 2160 × 1200, and a 90-Hz global refresh rate) at near (ca. 70 cm) or far (70 m) distances with same visual angle, which meant that the physical sizes of the stimuli at far distance were one hundred times bigger than those at near. The yellow fixation was presented at 2 m. And there was always a gray distance cue on the bottom of the screen (See Figure 1 for an illustration). Auditory stimulus was presented on the headphones of Oculus Rift.

Design and Procedure

A three-factor within-subjects design was used. The independent variables were distance, change, and sound. Distance referred to the memory array could present at near (0.7 m, 50% of the trials) or far distance (70 m, 50%). Change indicated that the test probe could be the same as the memorized target item (no change, 1/3 of the trials), differed only at the global level (global change, 1/3), or only at the local level (local change, 1/3). Sound had three conditions: 'absent' (1/3 of the trials), 'synchronous-target' (synchronous with the target which was going to be tested later, 1/3 of the trials), and 'synchronous-distractor' (synchronous with a distractor, one of the three letters which would not be tested later, 1/3 of the trials). Dependent variable was VWM performance, which was determined by a signal-detection-theoretic sensitivity measure: d-prime.

Each trial began with a yellow fixation lasted for 800 ms. Then the yellow fixation disappeared. After 800 ms, four Navon letters along with a white central fixation presented for 1200 ms at close (0.7 m) or far distance (70 m). Within 1200 ms, one randomly chosen element letter (small letter) within each Navon compound (large letter) changed color (from red to green or vice versa) continuously in every 50 ms (each will change 6 times in random order). In 1/3 of the blocks, sound was absent, while in 2/3 of the blocks, one letter changed its color with synchronous sound beeps. In half of the trials within sound-present blocks, beeps were synchronous with the target, and in the other half, beeps were synchronous with a distractor.

Participants were instructed to memorize all presented Navon letters at both global and local levels and were asked to fix on the central fixation. When the memory array was over, there was a blank screen for 1000 ms, followed by a test probe. When sound was absent, the test probe was presented at one, random position where memory array presented before. In the 'synchronous-target' condition, the test probe was presented at the position of the Navon letter which changed color with synchronous sound. And in the 'synchronous-distractor' condition, the test probe was presented at one of the other 3 positons where those Navon letters changed color without synchronous sound.

The task was to decide whether the test probe was the same (at both global and local levels) or different from the item that had been previously presented at the same location in the memory array. The probe item remained on-screen until a response was

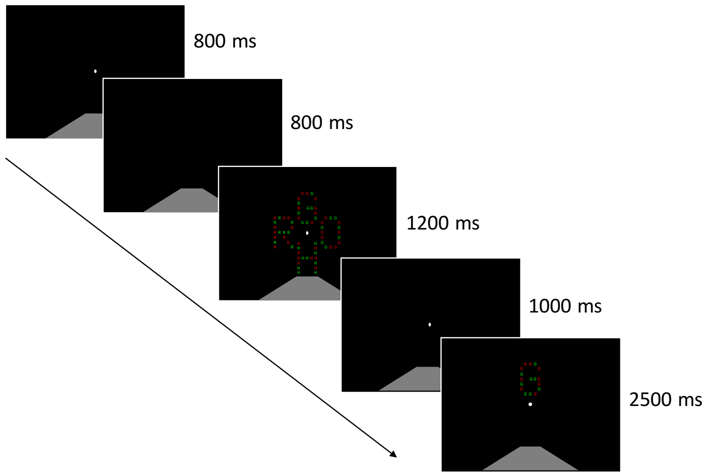


Fig. 2. An example trial of the experiment. After a yellow fixation and a brief interval, memory array was presented. Within 1200 ms, one local letter within each Navon letter changed color constantly in every 50 ms. In 2/3 of the blocks, one letter changed color with synchronous sound beeps. After a brief delay, a test probe was presented. It could be the same as the previous letter or differed at the global or local level.

given, otherwise it disappeared after 2500 ms. Participants were instructed to respond as accurately as possible. The inter-trial interval (ITI) was 500 ms. Figure 2 showed a trial sequence.

Change type was balanced and randomly mixed within 72 trials in each block. Participants received 2 sound-absent blocks and 4 sound-present blocks (presented in random order), and received feedback about their overall mean accuracy after each block. Before the formal experiment, 1 block of 36 practice trials was given to familiarize participants with the task. And they were provided with a green or red dot for 500 ms after each trial, indicating the answer was right or wrong, respectively.

Results and Discussion

Global Precedence: VWM Representation of Compound Stimuli

We performed a 2 (distance: near or far) × 2 (change: global or local) × 3 (sound: absent, synchronous-target, or synchronous-distractor) repeated-measures analysis of variance (ANOVA). The results were presented in Figure 3. The ANOVA revealed main effect of change to be significant: $F(1, 29) = 15.48, p < .001, \eta^2 = .348$. Global changes were detected more efficiently than local changes (mean precedence effect in d-prime: 0.82). The result showed an overall global precedence effect, which was consistent with Nie, Müller, and Conci's (2017). It indicated that VWM representations were organized in a hierarchically structured fashion reflecting the global/local structure of the perceptual inputs.

However, there was no significant interaction between change and the other two variables, which meant both at different distance and in different audiovisual environment, global precedence effect was stable. It was inconsistent with our expectation that, compared to far distance, near distance may facilitate processing of local letters according to CLT.

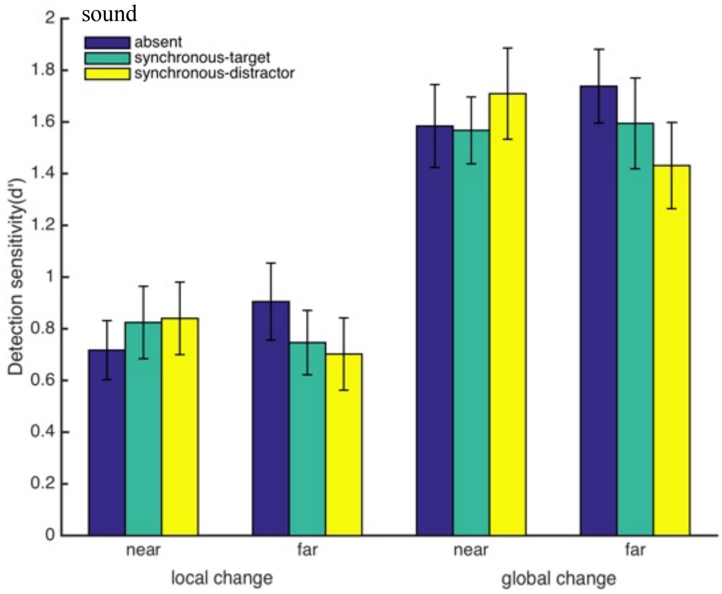


Fig. 3. Results. Mean d-primes as a function of distance and change, separately for 'absent', 'synchronous-target', and 'synchronous-distractor'. The error bars represent the standard errors.

We controlled the visual angle to be the same at near and far distance, given that it would affect global precedence effect (Kinchla & Wolfe, 1979). This could be the reason why distance did not play a modulating role in perceptual construal level, because one object ought to be seen as big when near and small when far. Hence, in our future experiment, we consider to present objects of the same physical size at different distances (i.e. with different visual angles), and also present objects of different physical size at the same distance (to make visual angles be the same as the former condition).

Asynchronous Sounds will Hamper VWM at far Distance

The two-way interaction of distance and sound was significant, $F(2, 58) = 3.78, p < .05, \eta^2 = .115$. At far distance, the effect of sound was significant, $F(2, 58) = 4.10, p = .022$. Change detection sensitivity was greater in 'absent' condition than 'synchronous-distractor' condition. At near distance, the effect of sound was not significant, $F(2, 58) = .75, p = .475$ (Figure 4).

After the experiment, some participants reported the sound beeps were a bit noisy and when sound was absent, they could concentrate more on visual task. The reason why sound had little effect on VWM at near distance could be that when visual objects were close, the visual signals were strong and the effect of sound was weak. As the distance increased, visual signals became weak and the effect of sound enhanced. The visual objects which was not synchronous with sound would easily be ignored and could hardly be recognized in the test. While the sound synchronous with the target might have a similar visual capture effect as pip and pop effect (an auditory pip made the synchronized visual object pop out from the environment, Van der Burg et al., 2008).

Nevertheless, the synchronous sound didn't show much advantage. The reason could lie in the visual objects and scene. The pip and pop effect was interpreted as a

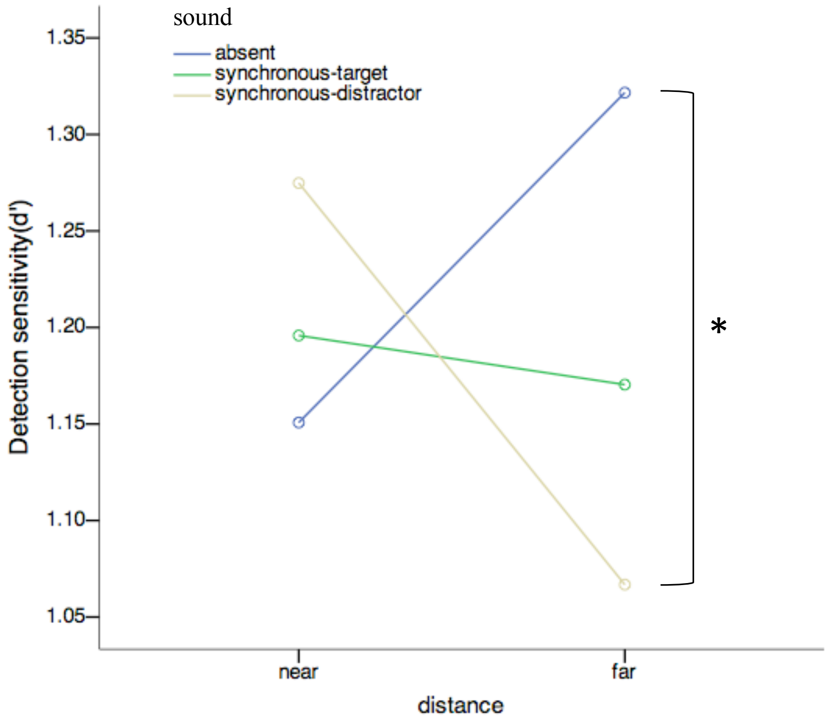


Fig. 4. Two-way interaction of distance and sound. The effect of sound was significant at far distance, but not at near distance (*: $p < .05$).

facilitation of audiovisual integration. The benefits of multisensory integration are proportionately greatest when cross-modal cues are weakest, and as the cues become more effective, unisensory component responses become more vigorous, and integrated responses become proportionately smaller (Stein & Stanford, 2008). In pip and pop experiment, the visual objects were simple lines and the scene was cluttered, which made visual target really weak and thus auditory cue stood out. While in our study, the visual stimuli were Navon letters forming gestalt structures and there was no visual distractor. Under such a circumstance where the visual information was robust, audiovisual integration did not take in effect. Given the possible issues, in the future experiment, we plan to make the gestalt structure of the memory array less conspicuous by presenting the four Navon letters at different orientations, along with several cluttered small letters in the background around them.

References

Bigelow, J., & Poremba, A. (2016). Audiovisual integration facilitates monkeys' short-term memory. *Animal Cognition,* 19(4), 799–811.

Hollingworth, A. (2004). Constructing visual representations of natural scenes: the roles of short- and long-term visual memory. *Journal of Experimental Psychology. Human Perception and Performance,* 30(3), 519–537.

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. Journal of Experimental Psychology: Learning Memory & Cognition, 26(3), 683–702.

Kinchla, R., & Wolfe, J. (1979). The order of visual processing: "Top-down," "bottom-up," or “middle-out.” *Perception & Psychophysics,* 25(3), 225–231.

Liberman, N., & Förster, J. (2009). The effect of psychological distance on perceptual level of construal. *Cognitive Science,* 33(7), 1330–1341.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognition,* 9, 353–383.

Nie, Q., Müller, H. J., & Conci, M. (2017). Hierarchical organization in visual working memory: From global ensemble to individual object structure. *Cognition,* 159, 85–96.

Schmidt, B. K., Vogel, E. K., Woodman, G. F., & Luck, S. J. (2002). Voluntary and automatic attentional control of visual working memory. *Perception & Psychophysics,* 64(5), 754–763.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews. Neuroscience,* 9(4), 255–66.

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology. Human Perception and Performance,* 34(5), 1053–1065.

Van der Stoep, N., Nijboer, T. C. W., Van der Stigchel, S., & Spence, C. (2015). Multisensory interactions in the depth plane in front and rear space: A review. *Neuropsychologia,* 70, 335–349.

Welch, R. B., & Warren, D. H. (1980). Immediate Perceptual Response to Intersensory Discrepancy. *Psychological Bulletin,* 88, 638–667.

Xie, Y., Xu, Y., Bian, C., & Li, M. (2017). Semantic congruent audiovisual integration during the encoding stage of working memory: an ERP and sLORETA study. *Scientific Reports,* 7(1), 5112.

LEARNING ARBITRARY ASSOCIATIONS BETWEEN VISUAL AND TACTILE FEATURES

Ryo Teraoka
Research Institute of Electrical Communication and Graduate School of Information Sciences, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan

Wataru Teramoto
Department of Psychology, Kumamoto University, 2-40-1 Kurokami, Chuo-ku, Kumamoto 860-8555, Japan

<terar@dc.tohoku.ac.jp, teramoto@kumamoto-u.ac.jp>

Abstract

Recent studies have demonstrated that a new association between auditory and visual features can be rapidly formed even in adults (Teramoto et al., 2010). This study investigated whether this finding can be generalized to the visuotactile combination. In a 9-min exposure session, two circles placed side by side were presented in alternation, and each onset was accompanied by a vibrotactile stimulation on two different positions of the forearm. We then quantified pre- and post-exposure perceptual changes using a motion nulling procedure. Results showed that after prolonged exposure to the visuotactile stimuli, the tactile sequence became a driver for visual motion perception, indicating that new associations can be formed between the visual and tactile modalities. Notably, the effect did not occur when the forearm posture was changed between the exposure and test phases, suggesting that the new representations were established after integrating proprioceptive information.

Our brain establishes the perception of an external world using information from various sensory modalities. For example, when we type on keyboard, we receive visual, auditory, and tactile signals. Such sensory signals are automatically and efficiently integrated to establish coherent and robust perceptions and cognition (Ernst and Bülthoff, 2004). One possible strategy to accomplish integration of such signals is to form associations between signals from different senses on the basis of past experience, such as the frequent occurrence of these signals. Indeed, it has recently been demonstrated that the brain rapidly forms an association between a sound sequence and visual motion when concurrently presented. Once this association has been formed, the sound sequence can be a driver for a visual motion perception. This phenomenon is called as "sound-contingent visual motion perception (SCVM; Teramoto et al., 2010; Hidaka et al., 2011, Kobayashi et al., 2012a, b)". In these studies, two circles placed side-by-side were alternatively presented, and each onset was accompanied by tone burst of a specific and unique frequency during the exposure phase. After exposure for a few minutes, the tones began to trigger illusory visual motion in response to a static visual object. These findings suggest that a new association between vision and auditory modalities and be utilized for visual motion perception.

A question for the present study is whether a new association regarding motion perception can be established between vision and touch as well, using a technique similar to SCVM. Cross-modal interactions in motion perception that have been reported thus far are not limited to the audiovisual domain. Bensmaïa et al. (2006) and Craig (2006) showed that the perceived direction and speed of tactile motion affected by visual motion.

Not only visual effects on tactile motion perception but also the reversed relationship has been reported. Blake et al. (2004) showed that ambiguous visual motion could be disambiguated by tactile information. However, it is not clear whether a similar association as observed in SCVM studies would arise for the visual and tactile modalities. Although it is not related to motion perception, Ernst (2007) demonstrated that one could learn to associate between features of the visual and tactile modalities (i.e., luminance and stiffness) that were previously unrelated, indicating that new associations can be formed in the visuotactile domain as well. In the present study, we investigate whether "touch-contingent visual motion perception (TCVM)" could be observed through exposure the onsets of the circle presentations were synchronized with vibrotactile stimulations on two different position of the forearm.

Furthermore, the present study also investigates which tactile reference frame is involved in the formation of this type of visuotactile association. Previous SCVM studies demonstrated that the selectivity of the contingent effects for the visual field (Teramoto et al., 2010), eye (Kobayashi et al., 2012a), sound frequency, and ear (Kobayashi et al., 2012b). These finding suggest that relatively early perceptual processing stages of the visual and auditory systems are involved in the production of SCVM. Regarding tactile processing, the location of tactile stimuli is initially anchored in a skin-based reference frame in early somatosensory processing (Penfield and Boldrey, 1937). Subsequently, it is anchored into more abstract reference frame such as body-centered or world-centered in later stages of processing (Graziano et al., 2004). To determine which of these tactile reference frames are involved in TCVM, the present study used a posture change manipulation. If TCVM is solely determined by the brain areas that have somatotopic maps, then it would occur even in the posture change condition.

Methods

Participants

Thirteen male volunteers, including the author (R.T), had normal or corrected-to-normal vision and normal hearing. The volunteers except for the author were naïve to the purpose of this experiment. Informed consent was obtained from each volunteer before concluding the experiment.

Stimuli

A white circle (2° in diameter, 5.2 cd/m^2) was presented as a visual stimulus on a black background on a 22-inch CRT display (60 Hz) with a viewing distance of 62 cm. A red bull's eye shape fixation (0.716° in diameter; 15.5 cd/m^2) was also presented either to 10° left or right of the center of the display. Tactile stimuli (50 ms, 300 Hz) were delivered by a pair of vibration stimulators with a contractor of a 2.0-cm diameter disk. The stimulators were strapped either to the participant's left or right ventral forearm, with one placed 5 cm from the center of the forearm to the wrist (hereafter, "Tact A") and the other 5 cm to the inside of the elbow (hereafter, "Tact B"). To mask any possible noise made by the vibrotactile stimulators, white noise was delivered through headphones throughout the experiment (79 dB SPL). Participant's head was stabilized by a chin rest.

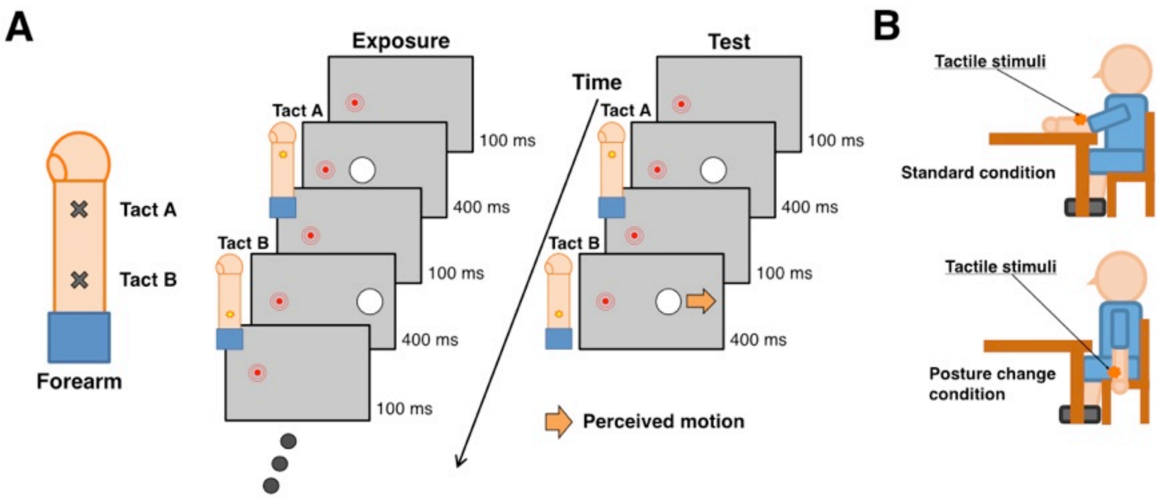


Fig. 1. Experimental design. (A) During the exposure phase, participants were exposed to visual motion in which two white circles placed side-by-side were alternately presented. The onset of the leftward circle was accompanied by a Tact A and the rightward circle a Tact B. During the test phases, a white circle was presented twice. The circle was perceived as lateral motion in the same direction as the previously presented visual motion when the onset of the circle was synchronized to the tactile sequence and the forearm posture was unchanged between the exposure and test phases. (B) The posture of the forearm in the exposure. The forearm was placed on the table, extended toward the display monitor in the pre-test, exposure, and all post-test phases ("Standard posture") except for the posture change test, where the forearm was off the table and extended to the floor ("Changed posture").

Procedure

The experiment consisted of 3 phases ¬: pre-test, exposure, and post-test. During the 9-min exposure phase, participants rested their arm to which tactile stimulators were strapped on the table perpendicular to the display monitor with their palms facing up. In this phase, two white circles were presented in alternation at 7.5° and 12.5° to the right of the fixation for half of the

participants and to the left of the fixation for the remaining participants. The duration of each circle was 400 ms and stimulus onset asynchrony (SOA) was 500 ms. For the half of participants, the onset of the leftward motion was synchronized to the tactile sequence of Tact B to Tact A (i.e., distal direction), while leftward motion was synchronized to the tactile sequence of Tact A to Tact B (i.e., proximal direction). The opposite pairing was employed for the remaining participants. The tactile sequence associated with leftward visual motion in the exposure phase was defined as "L-tactile" condition. Conversely, the tactile sequence associated with rightward visual motion was defined as "R-tactile" condition. Participants were instructed to focus on the fixation point during exposure phase without making any judgements and were allowed a short break every 3 min.

In the test phases, the strength of illusory motion was quantified by a motion-nulling procedure, wherein the physical visual displacement that corresponded to the point of subjective stationarity (PSS) was measured. A white circle was presented twice

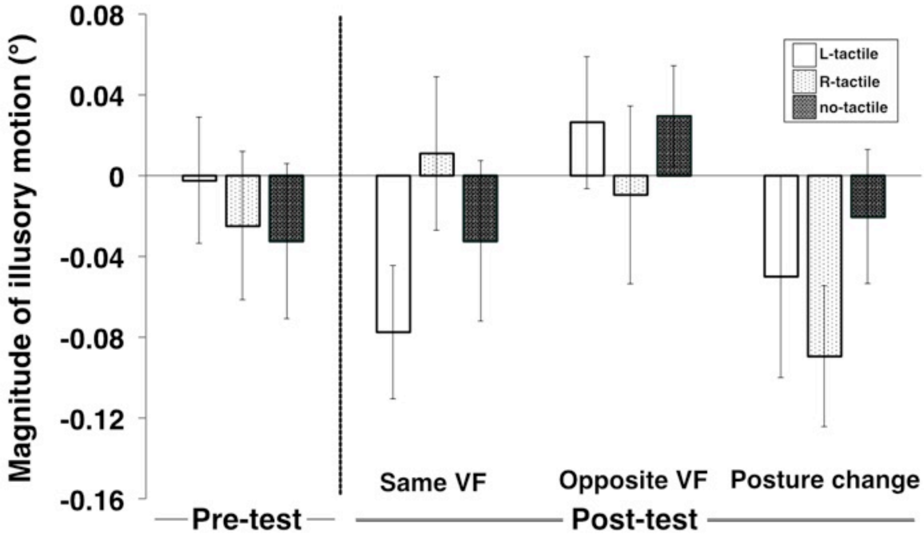


Fig. 2. Magnitude of illusory visual motion (°). Positive and negative values represent the illusory visual motion in the rightward and leftward directions, respectively. White bars represent the results for L-tactile condition, space-dots bars R-tactile condition, and dense-dots bars no-tactile condition. Error bar denotes the standard error of the mean.

with a 500 ms SOA, synchronized with two vibrotactile stimulations. The white circle was displaced by various angles from left to right or vice versa, from -0.8° to 0.8° in steps of 0.08° except for 0.0° (negative and positive values indicate leftward and rightward displacement angles, respectively), centering at 10° to the right (or left) of the fixation. The displacement angles were determined using the randomized maximum likelihood sequential procedure (Takeshima et al., 2001). After the test phase, a cumulative normal distribution function was fitted to all obtained responses and the PSS (a displacement angle at 50% point of this function) was estimated for each participant for each condition. There were three tactile conditions: R-tactile, L-tactile, and no-tactile conditions. The participants were asked to judge whether the visual stimulus moved leftward or rightward while ignoring tactile stimulation. Each tactile condition was repeated 80 times. The order of these conditions was randomized. In the pre-test, the participants' arm was placed in the same position as in the exposure phase.

The post-test comprised three sessions, which differed from each other in tested visual field and posture: no posture/ visual field change, visual field change, and posture change test.

During the no posture/ visual field change session, the stimulated visual field and forearm posture were the same as in the exposure. During the posture change session, the participant's arm hung vertically toward the floor. During the visual field session, the center of the white circle was placed in the opposite visual field in which the visual stimuli were presented in the exposure. The order of these post-test sessions was randomized among participants. The first post-test session began soon after the exposure and a 1-min re-adaptation was inserted before the second and third test sessions.

Results

Effects of the Associative Learning

One participant was removed as an outlier because the percentages of correct responses at the maximum visual displacement angle (0.8°) were below 75% in the no-tactile conditions, so that the fitting curve clearly deviated from those of the other participants. Data from the remaining 13 participants were analyzed. A two-way repeated measures analysis of variance (ANOVA) with phase (2; pre-/post-test) and tactile condition (3; L-/R-/no-tactile) showed the significant interaction effect between the factors ($F_{2,24} = 6.41, p = .006$). Regarding simple main effect of the tactile conditions ($F_{2,52} = 7.25, p = .003$), a post hoc test (Tukey's HSD) revealed that there were significant differences between all pairs of the tactile conditions ($p < .001$). In contrast, in the pre-test, no effect of the tactile condition was observed ($F_{2,52} = 1.38, p = .271$). These results revealed that the tactile positional change acquired driving effects for visual motion after a few minutes exposure, similar to the SCVM studies.

Selectivity of Visual Field

Previous SCVM studies showed that the association effects were observed only in the retinal position or visual field in which the audiovisual stimuli were presented during the exposure. The visual field change test in this study was included to clarify whether this selectivity was observed in TCVM as well. As mentioned above, in this test, visual stimuli were presented in the visual field where visual apparent motion was not presented during the exposure. A two-way repeated measures ANOVA with phase (2; pre-/post-test [visual field change]) and tactile condition (3; L-/R-/no-tactile) revealed no main and interaction effects, indicating that there is a visual field selectivity in TCVM. This finding also suggests that the observed TCVM cannot be explained only by the tactile-induced bias on the participant's judgement.

Effects of Proprioceptive

Here, the posture change test was analyzed. If the somatosensory processing stages in which tactile positions are coded in a somatotopic manner are crucially involved, TCVM should be observed in the posture change test as well. A two-way repeated measures ANOVA with phase (2; pre-/post-test [visual field change]) and tactile condition (3; L-/R-/no-tactile) was performed on the PSS data. No main and interaction effects were significant. This finding indicates that there is the posture selectivity in TCVM.

Discussion

The present study focused on the touch contingent visual motion perception (TCVM). After short-term exposure to visual apparent motion with vibrotactile stimulations on different position on the forearm, the tactile sequence affected visual motion perception. When the leftward visual apparent motion was paired with the distal vibrotactile sequence in the exposure phase, the tactile sequence induced greater leftward than rightward visual motion perception in the post-test. In contrast, when the rightward visual apparent motion was paired with the distal visuotactile sequence in the exposure phase, tactile sequence induced greater rightward than leftward visual motion perception in post-test. Furthermore, this effect did not occur when forearm posture was changed between the exposure and test phases, suggesting that the association is formed after integrating proprioceptive information.

Previous studies on SCVM have demonstrated that an association between vision and auditory modalities can be formed after a few minutes associative learning (i.e., exposure phase). Our current study clearly shows that the finding can be generalized to the visuotactile combination. Recently, several studies have reported that a sensory combination other than audition and vision could produce association effect, such as motor (Wallis and Backus, 2016) and olfactory (Kuang and Zhang, 2014). Thus, it is highly likely that the rapid formation of associations between signals from different modalities and its utilization for sensory binding are general rules for multisensory integration in the brain.

It should be noted that TCVM did not occur when the post-test was conducted with the arm in a different posture from that used in the exposure phase. These results suggest that brain areas that have a somatotopic map, which are mainly utilized in the relatively early stages of somatosensory processing, are not crucially involved in TCVM. In previous SCVM studies, there were the selectivity of retinal position (Teramoto et al., 2010), eye (Kobayashi et al., 2012a), ear, and sound frequency (Kobayashi et al., 2012b). These findings suggest that the brain can establish the associations in relatively early perceptual processing stages of the auditory and visual systems. Thus, the finding of the present study was not consistent with those of the previous studies.

One possible reason for this difference is the uniqueness of tactile processing. Previous studies reported that tactile stimulus location is automatically transformed from a somatotopic to external coordinate (Yamamoto and Kitazawa, 2001). Yamamoto and Kitazawa (2001) had their participants perform temporal order judgement about two tactile stimuli successively delivered to the tips of both hands. Although this task did not require, the crossing arms profoundly affected performance. Thus, such a characteristic of the tactile modality might explain the difference from the previous SCVM studies.

An alternative reason for this difference is that the level of processing involved is dependent on the stimulus feature used. While sound frequency was solely coded in the auditory modality (Kobayashi et al., 2012b), stimulus position can be coded in several modalities. Thus, for this type of association, the brain areas at which the relationship of stimulus features is more effectively encoded might be recruited. This hypothesis might be tested by using stimulus features unique to the tactile modality such as smoothness/roughness.

Acknowledgements

This research was supported by KAKENHI, Grant-in-Aid for Scientific Research (B) (No. 26285160) from the Japan Society for the Promotion of Science.

References

Azañón, E., Camacho, K., & Soto-Faraco., S. (2010) Tactile remapping beyond space. *Euro. J. Neurosci.,* 18, 1044–1049.

Bensmaïa, S. J., Killebrew, J. H., & Craig, J. C. (2008) Influence of visual motion on tactile motion perception. *J. Neurophysiol.,* 96, 1625–1637.

Blake, R., Sobel, K., & James, T. (2004) Neural synergy between kinetic vision and touch. *Psychol. Sci.,* 15, 397–402.

Craig, J. C. (2006) Vision motion interferes with tactile motion perception. *Perception,* 35, 351–367.

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *J. Vis.,* 7.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends. Cogn. Sci.,* 8, 162-169.

Graziano, M. S. A., Gross, C. G., Taylor, C. S. R., & Moore, T. (2004) A system of multimodal areas in the primate brain. In: C. Spence & J. Driver (Eds) *Crossmodal Space and Crossmodal attention,* Oxford University Press, 51–67.

Hidaka, S., Teramoto, W., Kobayashi, M., & Sugita, Y. (2011). Sound-contingent visual motion aftereffect. *BNC Neurosci.,* 12.

Kobayashi, M., Teramoto, W., Hidaka, S., & Sugita, Y. (2012a) Sound frequency and aural selectivity in sound-contingent visual motion aftereffect. *PloS ONE,* 7, e36803.

Kobayashi, M., Teramoto, W., Hidaka, S., & Sugita, Y. (2012b) Indiscriminable sounds determine the direction of visual motion. *Sci. Rep.,* 2.

Kuang, S., & Zhang, T. (2014) Smelling directions: olfaction modulates ambigious visual motion perception. *Sci. Rep.,* 4, 5796.

Penfield, W., & Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain,* 60, 389-443.

Takeshima, H., Suzuki, Y., Fujii, H., Kumagai, M., Ashihara, K., Fujumori, T., & Sone, T. (2001). Equal-loudness contours measured by the randomized maximum likelihood sequential procedure. *Acta. Acoust. United. Acoust.,* 87, 389-399.

Teramoto, W., Hidaka, S., & Sugita, Y. (2010). Sounds move a static visual object. *PLoS ONE,* 5, e12255.

Teramoto W, Kobayashi M, Hidaka S, Sugita Y (2013) Vision contingent auditory pitch aftereffect. *Exp. Brain Res.,* 229, 97–102.

Wallis, G. M., & Backus, B. T. (2016). When action conditions perception: Evidence of cross-modal cue recruitment. *J. Vis.,* 16, 6.

Yamamoto S, Kitazawa S (2001) Sensation at the tips of invisible tools. *Nat. Neurosci.,* 4, 979–980.

GEOMETRIC SHAPES MODULATE THE VISUAL ILLUSION INDUCED BY SOUND

Yasuhiro Takesima
Doshisha University, Kyoto, Japan
<yasuhiro.takeshima@gmail.com>

Abstract

Emotional information has an important function in sensory processing. Some previous studies have reported that simple geometric shapes also include emotional information; an inverted-triangle is processed similarly to an angry face. The present study examined the effects of emotional information on audio-visual integration using geometric shapes. I compared the occurrence rates of flash illusions among three geometric shapes: inverted-triangle, triangle, and circle. The flash illusions include fission and fusion illusions. Both illusions are induced by audio-visual integration. The experimental results indicated that the fission illusion was observed frequently in inverted-triangles as compared to triangles. On the other hand, the occurrence rates of the fusion illusion did not differ among the three types geometric shape. Each geometric shape, inverted-triangle, circle, and triangle, was rated as negative, positive, and neutral, respectively. Therefore, this study suggests that the fission illusion is modulated by geometric shapes that induce negative emotional information.

Multisensory integration is an important function for perception of the outer environment. Scientific interest in multisensory integration has increased recently (e.g., Stein, 2012), and many studies have examined process of integrating multisensory information, resulting in reports that multisensory percepts are stable and salient as compared to unisensory percepts (Stevenson et al., 2014, for a review).

However, the relationship between multisensory integration and emotional system has not yet been clarified. Emotional information is an important factor modulating sensory processing. Therefore, emotional information might affect multisensory integration. Some studies have examined the ability of emotional stimuli to modulate multisensory integration processes. Maiworm, Bellantoni, Spence, and Röder (2012) have shown that priming with a fearful voice reduces the ventriloquism effect (Jack & Thurlow, 1973). Kitamura, Watanabe, and Kitagawa (2016) have reported that task-irrelevant happy background music facilitated bounce percept in a stream/bounce display (Sekuler, Sekuler, & Lau, 1997).

On the other hand, the direct effects of emotional stimuli on multisensory integration have not been investigated sufficiently. In both Maiworm et al. and Kitamura et al., emotional stimuli were used as task-irrelevant stimuli. However, emotional stimuli themselves might directly affect multisensory integration processes if there are relationships between multisensory integration and the emotional system. Therefore, the present study investigated the effects of emotional information in flash illusions by using emotional stimuli as visual target stimuli.

Flash illusions are a type of famous illusion induced by audio-visual integration. When a brief single flash is presented with two simultaneous beeps, two flashes are often perceived; this phenomenon is called fission illusion (Shams, Kamitani, & Shimojo, 2002). In contrast, when two brief flashes are presented with one simultaneous beep, a single flash is often perceived, a phenomenon known as the fusion illusion (Andersen, Tiippana, & Sams, 2004). Both fission and fusion illusions are induced by audio-visual interactions. The fMRI studies have reported that activity of the primary visual cortex (V1) in the fission illusion is similar elicited by the presentation of two physical flashes (Watkins, Shams, Josephs, & Rees, 2006). Similarly, V1 activity in response to the fusion illusion did not differ significantly from the condition with just a single flash presented (Watkins, Shams, Tanaka, Haynes, & Rees, 2007).

The present study used simple geometric shapes as emotional stimuli. Fission and fusion illusions are induced by audio-visual interaction from occurring lower-order level sensory processing (Mishra, Martinez, & Hillyard, 2008; Mishra, Martinez, Sejnowski, & Hillyard, 2007). Moreover, both flash illusions are observed in a short temporal window. Therefore, visual stimuli of simple shapes would be valid for generating these flash illusions. Larson, Aronoff, and Stearns (2007) have reported that inverted-triangles are perceived with negative emotional valence. Furthermore, circles are rated positively (Larson et al., 2007). In the present study, inverted-triangle and circle are presented as negative and positive emotional stimuli, respectively. A triangle was used as neutral stimulus, as it has the same visual features as an inverted-triangle.

The purpose of present study was to examine how emotional stimuli directly affect visual flash illusion induced by sound. In the experiment, the occurrence rates of fission and fusion illusions were compared among the three simple geometric shapes. If the emotional system directly affects the multisensory integration process, the simple geometric shapes would modulate the occurrence rate of fission and/or fusion illusions.

Method

Participants

Eighteen observers, including the author (four women; mean age = 23.78 ± 2.72 years), participated in this experiment. All participants reported having normal or corrected-to-normal vision and normal audition. None had been informed as to the purpose of the experiment, except for the author. The participants provided written informed consent prior to the experiment.

Apparatus and Stimuli

Stimuli were generated and controlled by custom-made program written in MATLAB (The MathWorks Inc.), Psychtoolbox (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997), and a laptop computer (MacBook Pro, Apple; OS: macOS Sierra). The visual stimuli were displayed on a 17-inch CRT-display (Trinitron CPD-E230, Sony; resolution: 1024 × 768 pixels; refresh rate: 100 Hz). The auditory stimuli were conveyed through an audio interface (TAC-2R, ZOOM) and headphones (MDR-CD900ST, Sony). The simultaneity of the visual and auditory stimuli was confirmed through a digital oscilloscope (DS-5424A, Iwatsu). The experiment was conducted in a dim-dark room with 39.3 dB (A) of background noise. Participants viewed the monitor binocularly at a distance of 70 cm with their heads stabilized on a chin rest.

Each trial consisted of the presentation of a white fixation cross (0.5° of visual angle) and a visual stimulus; all stimuli were presented on a black background. The visual stimuli were three types of geometric shapes: triangle, inverted-triangle, and circle.

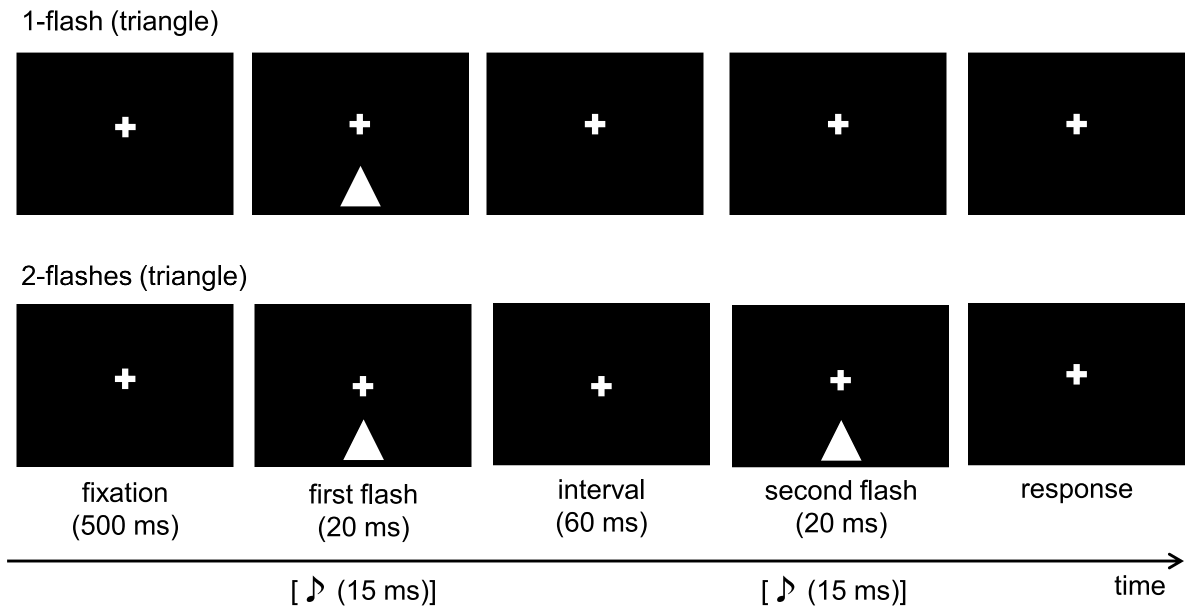


Fig. 1. Schematic representation of the experimental procedure. Top panels indicate the sequence of 1-flash. Bottom panels indicate the sequence of 2-flashes.

The size of each visual stimulus was within a 2.0° × 2.0° rectangle, the color was white, and the duration of presentation was 20 ms. The vertical distance between the fixation point and the center of the visual stimulus was 6.0 deg. The visual stimulus was presented once (1-flash) or twice (2-flashes) during a trial. The auditory stimulus was a pure tone at 1700 Hz. The duration of the auditory stimulus was 15 ms (including 2.5 ms at the start and end of the sound wave envelope), and the sound pressure level of the stimulus was 85 dB (A). The experimental condition for the auditory stimulus included three levels: no-beep, 1-beep, or 2-beeps. The no-beep condition had no beep sounds, the 1-beep condition and that a single beep during the first flash period, and the 2-beeps condition had beeps presented twice (during both the first and second flash periods). The stimulus onset asynchrony between the first and second stimulus presentations was 80 ms.

Procedure

A trial schematic is shown in Fig. 1. The trials were initiated by pressing the '0' key on a keyboard. Each trial consisted of a 500-ms fixation followed by the presentation of visual stimuli. The experiment followed a 3 (visual stimulus shape: shape) × 2 (number of flashes: flash) × 3 (number of beeps: beep) design. Each participant completed 360 trials (20 repetitions per condition). Participants were instructed to report the number of flashes they perceived by pressing one of two keys: '1' and '2' for one and two flashes, respectively.

After conducting this task, participants rated the simple geometric shapes used in this experiment. Participants were given a packet depicting one shape on each page. These geometric shapes were depicted in white within a black square (3 cm × 3 cm). Participants were asked to rate each shape in terms of bad–good, unpleasant–pleasant, unfriendly–friendly, and cruel–kind (see Lundqvist et al., 1994) using a 7-point bipolar semantic differential scale. Lower numbers reflected more negative ratings.

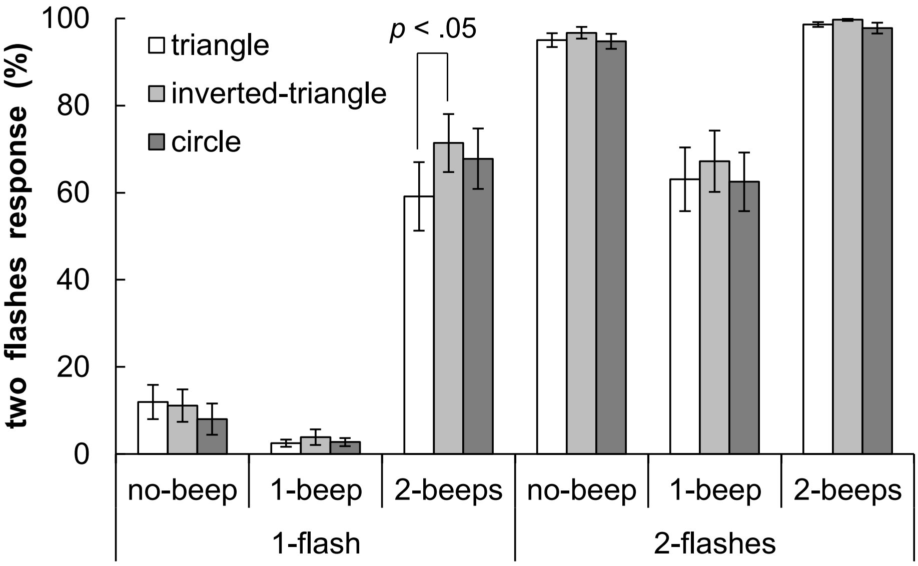


Fig. 2. Mean percentages of two flashes responses in each experimental condition. Error bars indicate the standard error of the mean ($n = 18$).

Results

First, the average subjective ratings combining the four semantic differential dimensions were calculated as an emotional valence score for each geometric shape. A one-way analysis of variance (ANOVA) with shape (3) as a within-participants factor was conducted. The main effect was significant [$F(2, 34) = 35.60, p < .001, \eta^2 = .68$]. Multiple comparisons (Shaffer's method) indicated that the inverted-triangle received lower scores ($M = 3.24, SD = 0.98$) than did the triangle ($M = 4.39, SD = 0.82; p < .01$). The circle ($M = 5.79, SD = 0.80$) received higher scores than did the triangle ($p < .001$).

The percentage of two flashes responses was calculated for each condition. The results are shown in Fig. 2. A three-way ANOVA with shape (3) × flash (2) × beep (3) as within-participants factors was conducted. The main effect of flash was significant [$F(1, 17) = 197.67, p < .001, {\eta_p}^2 = .92$], indicating that participants could discriminate correctly the number of flashes. Moreover, the three-way interaction was significant [$F(4, 68) = 2.82, p < .05, {\eta_p}^2 = .14$].

Further, two-way ANOVAs with shape (3) × beep (3) as within-factors were conducted for each flash presentation condition to investigate the effects of visual stimulus shape on the fission and fusion illusions. For the 1-flash condition, the main effect of beep was significant [$F(2, 34) = 87.89, p < .001, {\eta_p}^2 = .84$]. Multiple comparisons indicated that the percentages were higher in the 2-beeps than in no-beep and 1-beep conditions ($ps < .001$). Furthermore, the interaction between shape and beep was significant [$F(4, 68) = 4.25, p < .01, {\eta_p}^2 = .20$]. The simple main effect of shape was also significant in the 2-beeps condition [$F(2, 102) = 8.41, p < .001, {\eta_p}^2 = .14$]. Multiple comparisons indicated that the percentage was higher for the inverted-triangle than it was for the triangle conditions ($p < .05$). On the other hand, the percentages were not significantly different between the circle and triangle ($p = .33$).

For the 2-flashes condition, the main effect of beep was significant [$F(2, 34) = 26.32, p < .001, {\eta_p}^2 = .61$]. Multiple comparisons indicated that the percentage was lower in the 1-beep than in the no-beep and 2-beeps conditions ($ps < .001$). On the other hand, the interaction between shape and beep was not significant [$F(4, 68) = 0.27, p = .90, {\eta_p}^2 = .02$].

Discussion

The purpose of the present study was to examine the effects of emotional information in audio-visual integration. In this experiment, the occurrence rates of fission and fusion illusions were compared among three types of geometric shapes; triangle, inverted-triangle, and circle. Furthermore, the emotional valence of each geometric shape was investigated. The valence scores of the triangle, inverted-triangle, and circle were moderate, low, and high, respectively. Therefore, inverted-triangles and circles include negative and positive emotional valence, respectively, and while the triangle was a neutral stimulus. The experimental results indicated that two flashes responses were higher in the 2-beeps than in the no-beep condition when a single flash was presented. Moreover, two flashes responses were lower in the 1-beep than in the no-beep condition when two flashes were presented. Therefore, both fission and fusion illusions occurred in this experiment.

The occurrence rates of the fission illusion were modulated by the negative emotional information present in simple geometric shapes. The experimental results indicated that the occurrence rate of the fission illusion was higher in the inverted-triangle than it was in the triangle. The inverted-triangle was rated negatively, and therefore, the fission illusion was observed frequently in negative emotional stimuli as compared with neutral stimulus. Negative stimuli might enhance V1 activity in audio-visual integration. Negative stimuli induce fast amygdala activity, beginning 74 ms post-stimulus onset in facial expression study (Méndez-Bértolo et al., 2016). A direct pathway from the amygdala to V1 has been reported (Amaral, Behniea, & Kelly, 2003). On the other hand, auditory inputs also enhance V1 activity during the fission illusion (Watkins et al., 2006). Therefore, V1 activity is enhanced by negative emotional information, resulting in increases frequency of the fission illusion as compared with neutral stimuli. Alternatively, attention might be related to this different occurrence rate of the fission illusion between negative and neutral stimuli. The fission illusion is frequently observed at the attended visual field (Mishra, Martinez, & Hillyard, 2009). Negative stimuli capture rapidly attention compared with neutral stimuli (e.g., Most, Chun, & Widders, 2005). Therefore, it may be possible that attention explains the frequent occurrence of the fission illusion in the inverted-triangle.

On the other hand, the occurrence rates of fission illusion were almost same between positive and neutral stimuli. In the present study, the occurrence rate of fission illusion did not differ significantly between the circle and triangle. The circle was rated positively. Therefore, the fission illusion was not modulated by positive emotional stimuli. The facilitation of positive emotion in audio-visual integration has been shown previously (Kitamura et al., 2016). However, Kitamura et al. manipulated participants' emotional states, and thus, the direct effects of positive emotional stimuli on audio-visual integration were not examined. Moreover, the fast activity of the amygdala has not been observed for happy faces (Méndez-Bértolo et al., 2016). Therefore, positive emotional information might not affect audio-visual perception at lower-order levels of processing.

In contrast to the fission illusion, the occurrence rates of the fusion illusion were not modulated by emotional stimuli. The fusion illusion itself was observed, however the occurrence rates did not differ among the simple geometric shapes. Compared with the fission illusion, the fusion illusion is related strongly to individual differences in visual processing (e.g., Kostaki & Vatakis, 2016). Moreover, differences in processing mechanism between the fusion and fission illusions have been proposed (e.g., Kawaba, 2009). These differences might explain why negative emotional stimuli the fission illusion, but not the fusion illusion.

The present study found direct effects of emotional stimuli on audio-visual integration. A geometric shape rated as negative in emotional valence (inverted-triangle) facilitated illusory second flash perception. Therefore, this study found support for the relation between multisensory integration and the emotional system. However, further studies would need to investigate this relationship. For example, the temporal window was not compared between emotional and neutral stimuli. Moreover, the neural mechanism underlying the present effects has not been clarified. In this sense, future studies should build off the present work.

Acknowledgements

This work was supported by a Japan Society for the Promotion of Science KAKENHI Grant Number 15H06569 and 17K18069.

References

Andersen, T. S., Tiippana, K., & Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research,* 21, 301–308.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision,* 10, 443–446.

Jack, C. E., & Thurlow, W. (1973). Effects of degree of visual association and angle of displacement on the "ventriloquism effect." *Perception and Motor Skills,* 37, 967–979.

Kawabe, T. (2009). Audiovisual temporal capture underlies flash fusion. *Experimental Brain Research,* 198, 195–208.

Kitamura, M., Watanabe, K., & Kitagawa, N. (2016). Positive emotion facilitates audiovisual binding. *Frontiers in Integrative Neuroscience,* 9, Article 66.

Kleiner, M., Brainard, D., Pelli, D. (2007). "What's new in psychtoolbox-3?" Perception, 36, ECVP Abstract Supplement.

Kostaki, M., & Vatakis, A. (2016). Crossmodal binding rivalry: A "race" for integration between unequal sensory inputs. *Vision Research,* 127, 165–176.

Larson, C. L., Aronoff, J., & Stearns, J. (2007). The shape of threat: Simple geometric forms evoke rapid and sustained capture of attention. *Emotion,* 7, 526–534.

Maiworm, M., Bellantoni, M., Spence, C., & Röder, B. (2012). When emotional valence modulates audiovisual integration. *Attention, Perception, & Psychophysics,* 74, 1302–1311.

Médez-Bértolo, C., Moratti, S., Toledano, E., Lopez-Sora, F., Martínez-Alvarez, R., Mah, Y H., ⋯ & Strange, B. A. (2016). A fast pathway for fear in human amygdala. *Nature Neuroscience,* 19, 1041–1049.

Mishra, J., Martinez, A., & Hillyard, S. A. (2008). Cortical processes underlying sound-induced flash fusion. *Brain Research,* 1242, 102–115.

Mishra, J., Martinez, A., Sejnowski, T., & Hillyard, S. A. (2007). Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *The Journal of Neuroscience,* 27, 4120–4131.

Most, S. B., Chun, M. M., & Widders, D. M. (2005). Attentional rubbernecking: Cognitive control and personality in emotion-induced blindness. *Psychonomic Bulletin & Review,* 12, 654–661.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision, 10, 437–442. Sekuler, R., Sekuler, A., & Lau, R. (1997). Sound alters visual motion perception. *Nature, 385, 308.*

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research,* 14, 147–152.

Stein, B. E. (2012). Introduction. In Stein, B. E. (Ed.), *The New Handbook of Multisensory Processing.* London: The MIT Press; pp. xi–xvi.

Stevenson, R. A., Ghose, D., Fister, J. K., Sarko, D. K., Altieri, N A., Nidiffer, A. R., ⋯ & Wallace, M. T. (2014). Identifying and quantifying multisensory integration: A tutorial review. *Brain Topography,* 27, 707–730.

Watkins, S., Shams, L., Josephs, O., & Rees, G. (2007). Activity in human V1 follows multisensory perception. *NeuroImage,* 37, 572–578.

Watkins, S., Shams, L., Tanaka, S., Haynes, J.-D. & Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. *NeuroImage,* 31, 1247–1256.

CONFLICT MONITORING ENHANCEMENT WITH SHORT DURATION DEEP BREATHING: AN EVENT-RELATED POTENTIAL STUDY

Kok Suen Cheng[1], Yun Fah Chang[2], and Poh Foong Lee[1]*
[1]*Department of Mechatronics and Biomedical Engineering, University Tunku Abdul Rahman, Selangor, Malaysia*
[2]*Department of Mathematical & Actuarial Sciences, University Tunku Abdul Rahman, Selangor, Malaysia*
*Corresponding author: <leepf@utar.edu.my>

Abstract

Conflict monitoring and sustained attention, which forms part of the cognitive control, are essential to study the adaptability of actions in different situations. Mindfulness practices are reported to be able to increase the ability in conflict monitoring and sustained attention, but the various methods employed can be very time demanding. Looking at this issue, this study was focused on investigating the effects of deep breathing which constitutes a basic element in many mindfulness practices in increasing the sustained attention. Event-related potentials in the Go/NoGo paradigm were used to study the conflict monitoring process and sustained attention level immediately after a short section of deep breathing practice at 6 breaths per minute, as well as a follow-up section after practicing the deep breathing exercise for 7 days. The results showed that the short deep breathing practice was able to enhance the ability in conflict monitoring as seen from a greater NoGo N2 amplitude in the follow-up session, but not immediately after. For the sustained attention level, there were no changes as there was no behavioral nor Go P3 amplitude difference. This study showed that just by practicing the common element of deep breathing for 5 minutes every day it is possible to achieve at least one of the benefits of mindfulness practices.

The ability of humans to adapt their information processing and behaviors differently towards the current situation depending on the required goal is known as cognitive control. One mental information process entails this cognitive control is the conflict monitoring. Conflict occurs whenever there are two or more incompatible processes that tap into the same attentional resources at the same time (Botvinivk et al., 2001). In these situations, the incoming conflicting information that diverges from the norm need to be perceived and suitable adjustments to the behavioral action need to be made in order to monitor the problem. A cognitive task such as the Go/NoGo task has been established in assessing the conflict monitoring ability. Conflict occurs when overriding the prepotent action for the Go stimulus is needed whenever a NoGo stimulus appear. Furthermore, since the norm of this task is to respond to the majority Go stimuli, hence it gives space for the mind to wander around instead of being fully focused on the task, which in turn would drain the attention for carrying out the task and cause a decrement in the performance (Thomson, 2015). Thus, the Go/NoGo task has assessed the sustained attention for individual who do the test.

The event-related potential (ERP) is employed to study the specific brainwave pattern in responding to a particular event or stimulus as the ERP is excellent in probing the cognitive process due to its high temporal resolution. To integrate the cognitive task with ERP, ERP from the Go/NoGo task consists of two components that are being elicited by both the Go and NoGo stimuli, namely the N2 and P3 (Jodo & Kayama, 1991). The waveform from the NoGo trials is always being larger than that of the Go trials due to the NoGo task is N2 is related to the conflict monitoring (Donkers & Van Boxtel, 2004) and on the other hand, the NoGo P3 is related to the cancellation of the preplanned response towards the Go stimulus (Smith et al., 2007). Besides that, the Go P3 amplitude can serves as a marker for the sustained attention (Datta et al., 2007).

For brain train in increasing the sustained attention level together with a greater conflict monitoring ability, deep breathing exercise is proposed in this context. The basis of the deep breathing is to increase the mindfulness. Mindfulness is defined as the placement of one's attention to the current moment without any judgmental thought. According to Bishop et al. (2004), the mindfulness state of mind can be classified into two major facets: self-regulation of attention and orientating one's experience in the present moment. The first facet requires sustained attention to prevent the mind from wandering around, while the second facet includes the need of an awareness towards the conflicting judgmental thoughts. In accordance to the theory, research into mindfulness practices has shown that practitioners of such mindfulness have greater sustained attention and better conflict monitoring (for review see Chiesa et al., 2011). However, many types of exercises are being introduced in increasing the mindful level of an individual and not restricted to only one form of practice. Prior literatures have reported on attaining mindfulness with mindfulness practices as a whole package with all different styles of methods and total practicing time. The variance in these methods and the time demanding factor may form a potential barrier to pick up or maintain daily mindfulness practices (Carmody et al., 2008). In this study, the effect of video guided 5 minutes as short duration of deep breathing was used to investigate the conflict monitoring and sustained attention through the Go/NoGo task and present the significant changes in ERP waveform. It was hypothesized that deep breathing for a short duration can increase the conflict monitoring process as indexed by a larger N2 amplitude and the sustained attention as indexed by a larger P3 amplitude.

Method

EEG and ERP Data Acquisition

EEG signal was obtained using a NCC Medical 32 Channels Type A Routine EEG System (Model no: Nation 7128W-A32) with a sampling rate of 256 Hz. The raw data was processed using the FASTER (Nolan et al., 2010) whereby a statistical thresholding method with a Z score of $\pm$ 3 as the threshold value was used to remove artifacts in the channels, epochs, ICA and single-channel single-epochs. Furthermore, any epochs that have an amplitude greater than $\pm$ 75 μV were removed from subsequent analysis. The epochs range from -200 to 800 ms, with baseline correction from -200 to 0 ms. Separate ERP waveform was obtained for correct Go and NoGo trials. The N2 (from 100–300 ms) and P3 (from 300–600 ms) components were obtained from the averaged ERP waveform of the three midline electrodes (Fz, Cz and Pz). The peak latency and the area amplitude (integral area under or above the waveform) in the two time windows were extracted.

Subjects and Experimental Protocol

24 young adults (age: 20–27) were recruited and randomly split into 2 groups: Control (Con, $n = 13$), and Deep breathing for 5 minutes (DB5, $n = 11$). Exclusion criteria include: 1) those who were sick for the past two weeks prior the experiment, 2) having any medications, and 3) those who face difficulties in deep breathing for 5 minutes. All participants have normal or corrected-to-normal vision, no respiratory diseases or psychiatric disorders as reported by themselves. The research procedures were approved by the local university scientific and ethical review committee (Reference no.: U/SERC/04/2017). The participants understood the experimental procedures and an informed consent was obtained prior to the experiment.

The breathing rate for the deep breathing group was set to 6 breaths per minute. The DB group was requested to follow the instructions as shown in a video to achieve this breathing rate for each respective breathing duration of 5 minutes. In the video, there were appearing petals for the inhalation and disappearing petals for the exhalation process as shown in Fig. 1(b), with both process lasting for 5 seconds to achieve a rate of 6 breaths per minute. The participants were instructed to keep their focus on the video and also on their breathing, and also to feel the air going in and out of their body.

For controlled environment, the experiment was carried out in a room with natural sunlight brightness through glass window panel. Upon arrival to the laboratory, the participants were first instructed to rest for 15 minutes in a comfortable sitting position. At the same time, a NCC Medical 32 Channels Type A Routine EEG System (Model no: Nation 7128W-A32) was set onto the participants. A baseline reading was taken in the next 5 minutes (R1). Immediately after the baseline reading, the participants performed the Go/NoGo task (see Fig. 1(a)), which last about 10 minutes (T1). After finishing the test, the DB participants began the deep breathing intervention (INT). For the control group, they were instructed to relax and no instructions on deep breathing were given. After the intervention, the participants were requested to rest for another 5 minutes (R2) before performing the same Go/No-go task for a second time (T2). This concluded the first session.

For the undergoing deep breathing participants, they were required to perform the deep breathing intervention everyday once at any time suitable for them. The deep breathing video clip was sent to the DB participants and a reminder for them to practice in this 7 days were monitored through Whatsapp. After finishing the deep breathing exercise, the participants sent a confirmation to the investigators. As for the control group, no such instructions were given. All of the participants returned exactly one week after their first session at the same time slot for another follow-up section for the experiment. In the follow-up session, the baseline reading was again recorded for 5 minutes (R3). However, in this session, the Go/No-go task (T3) was performed once after the baseline reading.

Statistical Analysis

The behavioral results from the Go/NoGo task include the overall clicking accuracy, Go reaction time, omission error, commission error and reaction time variability. A 2×2 repeated measure ANCOVA with Group (Con and DB5) as the between subject factor while the Time (T2 and T3) is the within subject factor and the baseline reading at T1 as the covariate. For the ERP analysis, the amplitude and the latency of the Go and NoGo waveform were analyzed separately using a similar repeated measure ANCOVA as above.

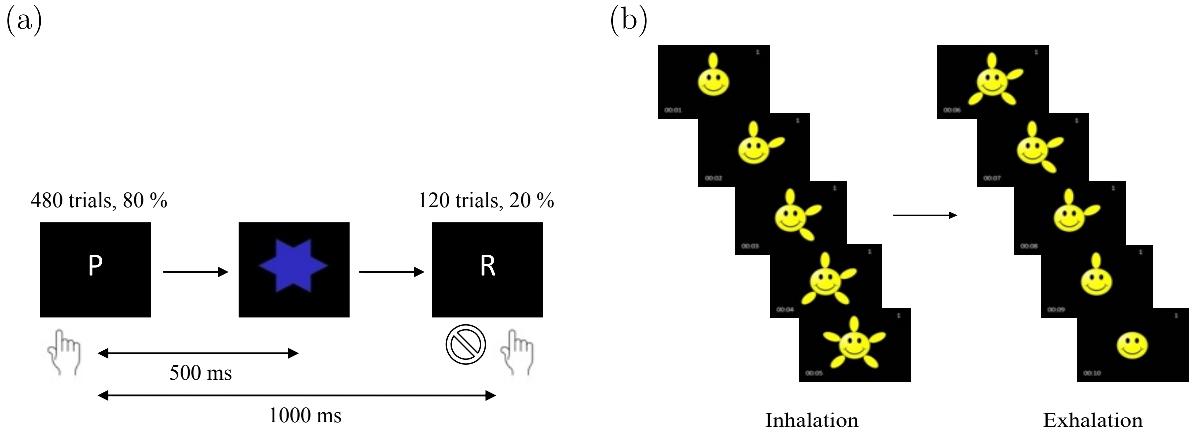


Fig. 1. (a) Stimuli in the Go/NoGo task whereby 'P's need to be responded and 'R's do not need to be responded. The stimulus time was 500 ms while the interstimulus time was 1000 ms. Trials were presented randomly in two blocks with a short break between them. (b) Screenshots of the guided deep breathing intervention video showing both inhalation and exhalation.

A p value of < 0.05 was considered as significant. The values are reported in mean and standard error in parenthesis.

Results

Behavioral Results

For the reaction time, there was a significant Time main effect $[F(1, 23) = 5.416, p = 0.029, \eta_p^2 = 0.191]$ indicating that the reaction time was shorter at T3.

ERP Result

The grand average waveforms for Go and NoGo trials at T1, T2 and T3 are shown in Fig. 2(a). For the Go N2 amplitude and latency, there were no main effect of Group, Time or interaction between them. For the NoGo N2 amplitude shown in Fig. 2(b), there was a significant Group main effect $[F(1, 21) = 7.543, p = 0.012, \eta_p^2 = 0.264]$, Time main effect $[F(1, 22) = 6.583, p = 0.018, \eta_p^2 = 0.230]$ and a significant Group $\times$ Time interaction $[F(1, 21) = 4.589, p = 0.044, \eta_p^2 = 0.179]$. Post-hoc analysis showed that at T3, the NoGo N2 amplitude of DB5 was larger than Con [0.167 (0.025) vs 0.070 (0.023), $p = 0.009$] and for DB5, the amplitude at T3 was larger than that of T2 [0.167 (0.025) vs 0.081 (0.015), $p = 0.004$)]. For the NoGo N2 latency, Go P3 amplitude, Go P3 latency and NoGo P3 amplitude, there was no significant main effect of Group, Time or interaction. However, there was a Group main effect approaching significance $[F(1, 21) = 3.973, p = 0.059, \eta_p^2 = 0.159]$ for the NoGo P3 latency with the latency of the DB5 shorter than that of Con as seen in Fig. 2(b).

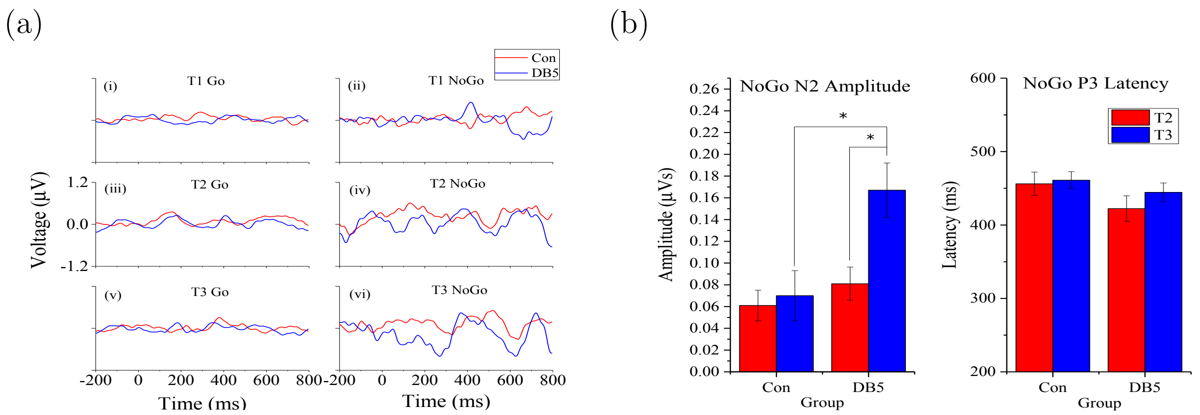


Fig. 2. (a) Grand average waveform for (i) T1 Go, (ii) T1 NoGo, (iii) T2 Go, (iv) T2 NoGo, (iv) T2 NoGo, (v) T3 Go and (vi) T3 NoGo. (b) The NoGo N2 amplitude and (c) NoGo P3 latency for both groups at T2 and T3. Error bars represent standard errors. *: $p < 0.001$.

Discussion

This study showed that performing a short 5 minute deep breathing at 6 breaths per minute everyday for a week was able to produce a difference in the neurophysiological measurement, but not at the behavioral level, as compared to a control group in the follow-up session but not immediately after. The increase in the NoGo N2 amplitude of the DB5 group implies that the deep breathing participants have enhanced conflict monitoring towards the NoGo stimulus by placing more attentional resources on it.

The guided deep breathing in this study mimics one of the exercise in mindfulness practices, which is the mindful breathing. In mindful breathing, the attention is placed onto the breathing exercise and when the mind starts to wander around, it is required to notice the wandering mind which is able to bring the attention back to the breathing exercise. The deep breathing in this study is guided by a video and this allowed the participants to focus on their breathing easily rather than having the deep breathing on themselves. The deep breathing guided video may be able to help the participants who has no prior experience in mindfulness training in having their mind at the present moment without being judgmental. Following this, literatures that report mindfulness practices that encompass the mindful breathing supports the

result obtained in the current study. A study by Tang et al. (2007) investigated the integrative body-mind training (IMBT) which includes body relaxation, breath adjustment, mental imagery and mindfulness training on 40 undergraduate students showed that the experimental group have enhanced conflict monitoring as indexed by a better performance in the Attentional Network Task (ANT) after five days of training. On the other hand, Menezes et al. (2013) reported that by paying attention to their breathing and counting the number of exhalation it was able to reduce the omission error in a visual discrimination task in the post-test session after two weeks of training. The result from this study extends the possibility of achieving an enhanced conflict monitoring just by performing deep breathing at 6 breaths per minute everyday as opposed to the whole set of mindfulness practices.

The current result is further supported by brain activation studies in mindfulness practices. The neural source of the NoGo N2 component is found to be at the anterior cingulate cortex (ACC) which is in line with the conflict monitoring interpretation (Nieuwenhuis & Yeung, 2003). Literatures studying the effect of focused meditation on the brain activation showed a greater activation of the ACC along with several frontal and parietal regions (Dickenson et al., 2013; Holzel et al., 2007). Collectively, this result implied that by focusing on the breathing (with the possibility of reducing the breathing rate) it is able to modulate the early awareness of distracting stimuli.

The lack of difference in the behavioral result, especially the omission error and reaction time variability and the Go P3 amplitude suggested that deep breathing at 6 breaths per minute might bring no effect to the sustained attention level. There is, albeit not statistically significant (p = 0.056), a trend showing that deep breathing can also increase the cognitive processing speed in canceling response towards the Go task. Further study with a greater sample size and more focused age range would be better to confirm this effect.

Conclusion

In this study, the effects of a particular component of mindfulness practices, namely the deep breathing on the conflict monitoring and sustained attention were investigated through a Go/NoGo paradigm and ERP. By deep breathing at a rate of 6 breaths per minute for a 5 minutes duration guided with a video, there was an increase in the NoGo N2 amplitude in the follow-up session as compared to the control group when performing a Go/NoGo task, which indicates an enhanced conflict monitoring ability. There was no effect on the sustained attention level as shown by the same Go P3 amplitude across groups. This study shows the possibility of performing only a short duration deep breathing as contrast to the whole mindfulness practice in order to reap the same benefits.

Acknowledgements

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. We would like to thank those who had participated in the experiment.

References

Bishop, S.R., Lau, M., Shapiro, S., Carlson, L., Anderson, N.D., Carmody, J., Segal, Z.V., Abbey, S., Speca, M., Velting, D., & Devins, G. (2004). Mindfulness: A proposed operational definition. *Clinical Psychology: Science and Practice,* 11, 230-241.

Botvinivk, M.M., Braver, T.S., Barch, D.M., Carter, C.S., & Cohen, J.D. (2001). Conflict monitoring and cognitive control. *Psychological Review,* 108, 624-652.

Carmody, J., & Baer, R. A. (2008). Mindfulness, spirituality, and health-related symptoms. *Journal of Psychosomatic Research,* 64, 393-403.

Chiesa, A., Calati, R., & Serrentti, A. (2011). Does mindfulness training improve cognitive abilities? A systematic review of neuropsychological findings. *Clinical Psychology Review,* 31, 449-464.

Datta, A., Cusack, R., Hawkins, K., Heutink, J., Rorden, C., Roberston, I.H., & Manly, T. (2007). *Computational Intelligence and Neuroscience,* 2007.

Dickenson, J., Berkman, E.T., Arch, J., & Lieberman, M.D. (2013). Neural correlates of focused attention during a brief mindfulness induction. *Social Cognitive and Affective Neuroscience,* 8, 40-47.

Donkers, F., & Van Boxtel, G. (2004). The N2 in go/no-go tasks reflects conflict monitoring not response inhibition. *Brain and Cognition,* 56, 165-176.

Holzel, B.K., Ott, U., Hempel, H., Hackl, A., Wolf, K., Stark, R., & Vaitl, D. (2007). Differential engagement of anterior cingulate and adjacent medial frontal cortex in adept meditators and non-meditators. *Neuroscience Letters,* 421, 16-21.

Jodo, E., & Kayama, Y. (1991). Relation of a negative ERP component to response inhibition in a Go/No-go task. *Electroencephalography and Clinical Neurophysiology,* 82, 477-482.

Menezes, C.B., de Paula Couto, M.C., Buratto, L.G., Erthal, F., Pereira, M.G., & Bizarro, L. (2013). The improvement of emotion and attention regulation after a 6-week training of focused meditation: A randomized controlled trial. *Evidence-Based Complementary and Alternative Medicine,* 2013.

Nieuwenhuis, S., & Yeung, Nick. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task: Effects of response conflict and trial type frequency. *Cognitive, Affective, & Behavioral Neuroscience,* 3, 17-26.

Nolan, H., Whelan, R., & Reilly, R.B. (2010). FASTER: Fully automated statistical thresholding for EEG artifact rejection. *Journal of Neuroscience Methods,* 192, 152-162.

Smith, J.L., Johnstone, S.J., & Barry, R.J. (2007). Response priming in the Go/NoGo task: The N2 reflects neither inhibition nor conflict. *Clinical Neurophysiology,* 118, 343-355.

Tang, Y., Ma, Y., Wang, J., Fan, Y., Feng, S., Lu, Q., Yu, Q., Sui, D., Rothbart, M.K., Fan, M., & Posner, M.I. (2007). Short-term meditation training improves attention and self-regulation. *PNAS,* 104, 17152-17156.

Thomson, D.R., Besner, D., & Smilek, D. (2015). A resource-control account of sustained attention: Evidence from mind-wandering and vigilance paradigms. *Perspectives on Psychological Science,* 10, 82-96.

EYE MOVEMENTS AND LIE DETECTION

Yulia V. Bessonova
Institute of Psychology, Russian Academy of Sciences, Moscow, Russia
Svetlana Pezchikova
Kutafin Moscow State Law University, Moscow, Russia
<farandi@mail.ru>

Abstract

Accurate value-driven attention and lie detection is crucial for psychodiagnostics validity. Eye-tracking is one way to improve it. The purpose of study was to find a nonverbal and verbal markers of deception. The first method consisted in stimuli' presenting with true or false answer; the second method was a combined analysis of verbal and oculomotor activity during 10 projective test images presenting (duration: 10000 ms) with 3 types of instructions: free reply, socially desirable reply, a hiding some elements of content. We found a complex of oculomotor markers for reliably lie detection which are stimulus and deception' type free. Specific oculomotor and verbal markers were established what allow differentiate any kinds of deception - random choice, socially desirable reply, hiding information. Special software has been developed based that results that identifying areas of value-driven attention and true answers, even respondents try to hide it in testing.

The polygraph testing technique remains a widespread instrumental method for lie detection despite gaining a lot of criticism. There have been critiques of the polygraph techniques from several perspectives. Arguments include that it based on a number of outstanding issues, including, first of all, the low validity of polygraph testing; the high variability of individual responses; the subjectivity of the polygraph examiner affecting testing process and test interpreting, and these cannot be eradicated; the differential impact of attitudes, described in detail by P. Ekman (1992); the complexity of testing in an altered state of consciousness, for example, a generalized fear, mood disorders, any form of intoxication etc. (Ioffe et al., 2007); a widespread resistance to testing, as well as difficulties in defining reasons can be given for psycho-physiological changes (the observed reaction is caused by exam situation or significant stimulus?). Moreover, in 2012, Jarrett noted that the scientific community does not have any theoretical evidence to substantiate the reliability of polygraph testing, which would be published in serious peer-reviewed sources (Jarrett, 2012).

These shortcomings call for either increasing validity and reliability of polygraph test or searching for other techniques of lie detection. The action within the scope of the first approach are geared towards the combination of verbal, non-verbal and physiological markers of deception (Vrij, 2008) implemented together with improved technology of questioning (Walsh & Bull, 2015). In the second area, the researchers concentrated on seeking another psychophysiological correlates of deception that would not be volitionally controlled, but rapidly changing and useful for differentiating between the changes owing to the functional state of the individuals or intention to hide true information. The eye movements are one of one of the most promising psychophysiological markers of deception.

Because of recent advances in technology, the eye-tracking has come to be utilized for psychological and psychophysiological research more frequently in recent years. Eye movements as markers of deception have several advantages. Firstly, the eye-movement registration demands little in terms of examiner's skill requirements. Secondly, the eye-tracking technology allows remote, non-invasive diagnostics, while at the same time delivering extremely precise results with less time spent calibration, testing and data processing. Thirdly, many psychophysiological indices are too often inert responding with great latency and, for that reason, uninformative. While eye-tracking has high- frequency data recording that enable us to capture more thin changes in eye movements , such as, saccade velocity etc. Fourthly, tracking is based on the measurement of cognitive responses instead of an emotional response that is considered unreliable, dependent on the respondent's status and testing situations. The main idea of eye-tracking is the "value-driven attention" - the searching and focusing more on a subjective significant and/or emotional stimulus. Therefore, the eye-tracking hasn't limitations of polygraph and makes it technically possible to differentiate true and false answers regardless of the emotion reaction of the testing situation, the functional state of the respondent, etc. The eye-tracking as perfect substitute for polygraph are implemented in investigating crimes by the police (Cook et al., 2012; Naganuma & Ochi, 2016). University of Utah professors and a Utah company have developed a new type of lie detector under the name of EyeDetect based on eye-movements.

The experience with the eye-tracking and knowledge gained in the studies confirm the validity of the eye-tracking lie detection, but data are contradictory and subject to varying interpretations depending on testing conditions. For example, the length of scanpass as a marker of lie depends on the age (Serras Pereira et al., 2016). The positive stimulus has a larger number of fixations and longer-term duration instead of the negative stimulus (Scott, 2009; Cook et al, 2012).

Some inconsistencies have been observed both in the set of eye movements as markers of lie, and their degree of shift, which makes it possible to distinguish between true and false answers. Thus, the purpose of this study is to differentiate true and false answers in various testing situations and to identify constant set of eye movement markers of deception.

Method

The experiment seems to reflect a view of three main types of lying including a concealment, a passive lying in terms of P. Ekman (1992), and an active deception, in turn, takes two broad forms (Vieira & Lane, 2013): "false descriptions" (an imaginary event) and "false denial" (a negation of a real event).

The procedure of the study has three phases: (a) the background phase was aimed at searching for an individual response rate and a selection of the specific eye movement markers of truth / falsity. Ten calibration questions were presented before the start of testing and involving reliable answers to standard socio-demographic questions (gender, age, education, etc.). After that, the same questions were presented again with instructions to knowingly give any false answer, (b) the second area of research is comparing eye movements response to text or image stimulus yielded with a view to identifying meaningful patterns. The pictures were previously selected by experts (10 people) as emotionally neutral. According to the theory of P. Kavanakh, these pictures were a closed circuit, black on a white background to reduce the cognitive processing complexity. The images were presented with two types of instruction: true answer and click on the correct picture (16 slides); false answer and click on the wrong picture (16 slides), (c) the third phase combined analysis of eye movements and verbal actions in case of true or false answer.

Stimuli - 10 projective pictures (Social Motivation Test, by Almaev & Murasheva), presented with three types of instruction: freely viewing and freely projective story; a socially desirable story; a story with partially concealed content.

The location of the pictures on the slide was subjected to a positional equalization. Time of presentation: 10000 ms. Presentation of stimuli was separated by a mask (25% gray color, duration: 500 ms).

The SMI iView Red 250Hz eye tracker was used to record the eye movements.

Respondents: N=108, mean age = 27, all of them employed full-time, 52% female and 48% male.

Results and Discussion

Eye Movements Lie Markers According to the Type of Stimulus

The discriminant analysis of eye movements in case of a true and false answers to text or image stimuli had provided valuable evidence for revealing a complex of informative markers. These markers are essentially the same and authentically identify a false answer irrespective of stimulus' type.

Regardless of the stimulus material (text or image), a similar set of interrelated eye movements markers of a false answer was identified. The true answer is accompanied by a predominance of focal attention: it's characterized by fewer repeated fixations and revisits, slower motor response, greater amount of fixation on the selected stimulus, combined with high-speed low-amplitude saccades.

Table 1. Eye movements' markers of true and false answers.

Eye movements parameters	Text stimulus True	False	p-level	Image stimulus True	False	p-level
Average Fixation Duration (ms)	411.78	344.02	0.000010	277.77	284.60	0.012199
Total Fixation Duration (ms)	1565.88	985.22	0.000000	16238.61	14643.52	0.014565
Fixation Count	4.18	3.08	0.002237	44.19	40.75	0.036894
First Fixation Duration (ms)	183.79	186.94	0.000000	228.64	226.11	0.014742
Saccade Count		no differences		45.35	42.19	0.047710
Saccade Amp. (°)	56.91	58.16	0.000145	5.76	7.45	0.044094
Saccade Velocity Average (°/s)	96.94	99.90	0.000003	111.23	128.21	0.043987
Saccade Velocity Peak (°/s)	180.44	186.55	0.021107		no differences	
Dwell Time (ms)	1620.52	1027.16	0.000374	1527.17	1504.39	0.018698
Glances Count	2.52	1.88	0.006091	2.01	1.90	0.029720
Revisits	2.67	4.77	0.019976	21.96	27.59	0.051900
Time to First Mouse Click (ms)	3586.37	2854.98	0.001830	539.71	522.32	0.039038

Note: only significant differences are included.

Table 2. Eye movements' markers of different types of deception.

Eye movements parameters	Type of deception			p-level
	Random choice	Socially desirable	Concealment	
Total Duration Blink	504154.46	336924.85	307088.52	0.000000
Fixation Count	98.15	106.25	90.07	0.000000
Fixation Duration	197125.80	227438.93	217156.76	0.000000
Fixation Average Pupil Diameter (mm)	4.12	4.04	4.03	0.000000
Saccade Count	108.06	113.52	93.34	0.000000
Saccade Total Duration	25875.47	22491.72	24066.30	0.000000
Saccade Amplitude Average (°)	4.77	3.71	3.60	0.002733
Saccade Velocity Peak (°/s)	303.30	292.34	256.99	0.000000
Saccade Velocity Average (°/s)	136.35	123.04	107.68	0.000000
Saccade Acceleration Peak (°/s)	11270.56	9555.72	8904.75	0.000000
Revisits	33.97	38.95	39.82	0.051900
Dwell Time (ms)	1527.17	1504.39	1425.27	0.018698
First Fixation Duration (ms)	282.32	200.05	185.81	0.014742
Average Fixation Duration (ms)	329.50	311.48	305.95	0.012199
Glances Count	2.51	1.65	1.62	0.029720

Selection of false answer is manifested in a change in the sequence of viewing, increasing amplitude and speed of saccades, growth in the number of repeated fixations with a decline in the number of fixations on selected false answer and less duration of fixation on it, decreasing in the total time of reviewing all slide. The choice of a false answer, especially with the number of alternatives more than two, reflects the process of an additional cognitive task, which is consistent with the findings of a study conducted by A. Cook et. al. (2012).

Eye Movements and Different Types of Deception

Several eye movement parameters depend on the type of deception. The random choice of false answers (equivalent to the passive lying) is accompanied by the predominance of the orienting viewing (lengthening the overall gaze trajectory, growth in the number of revisits, the greater amplitude and velocity of the saccades) and accelerating the motor response. A random choice does not correspond to the previously identified "value-driven" eye-movement markers, that can be seen as a measure of "passive lying" without additional cognitive burden.

The types of active lying are marked by a sharp decrease in the number of blinking and narrowing of the pupil. In addition to the general criteria for deception, the choice of answer in the mould of the false descriptions (equivalent to the socially desirable story) has also been accompanied by a rise in the number and duration of fixations on the selected false stimulus, increasing number of saccades and decreasing in saccadic speed parameters, which means the incorporation of the selected stimulus in the context of false verbal answer and it thus becomes important.

The false denial answer (a story with partially concealed content) differs over sharp has decrease in the total time of reviewing, the total number of fixations, as well as decrease in both quantitative and speed indicators of saccades, but the number of revisits has increased. The analysis of the pattern of viewing carries information about areas of the stimulus field that have subjectively value. In case of image stimulus, the time and speed rates of eye movements are increasingly important than quantitative rates. Quantitative measures (the number of revisits, fixations, saccades, etc.) reflect the characteristics of viewing and depend on complexity of image, as well as the average duration of the first fixation. When the density of the stimulus field is increased, they could not be regarded as markers of deception, because active lying integrates the secondary objective to, in the process, construct a story with inclusion and hidden some context.

Eye Movements and Verbal Markers of Lying

An additional markers of deception were revealed in the jointly analysis of verbal activity and eye movements that performed the “redundancy” functionality. The instructions provoking deception cause general changes in speech: verbalization latency increases to 2.5-3 sec. in parallel with the increase in the number of word repetition and/or a common narrative in other words; a latency between viewing the picture and inclusion it in the story is growing. A motor response is accelerating despite the increase in the numbers of the erroneous choices, and the motor reaction prevailed over the verbal - first, the respondent selects a false answer, then voiced it. In case of the positive information is hidden, the latency of the verbalization decreases, although timing of the mouse click has remained constant. In case of the negative information is hidden, the latency of the verbalization increases, but the timing of mouse click on the selected false answer is shortened.

The stereotype storylines (narrative patterns limiting) are the common to all participants feature of any kind of deception. Separately, the time of latency before beginning of verbalization is no longer appropriately characterized as one of deception sign, because some respondents were afraid of the exam situation and another have a problem with verbalization or didn’t have the necessary story-telling skills.

A specific features of different types of deception have been found. In the case of concealment, the prevailing motor reaction over the verbal could be considered an additional markers as well as the shortening the story. The instruction caused the concealment contributes to the reduction of speech production (from 139 words to 102 words in concealment situation), but it does not automatically reduce the number of fixations on the hiding stimulus. Instead, the instructions caused the social desirably story is leading to extension of story and to increase of verbal production to an average of 173 words.

A feature of socially desirable narrative is a “separation from context” (fantasy without reliance on source stimulus). Another feature of socially desirable story is the manifestation of a personal feelings about the depicted situation or characters, fantasizing about their motivation and desires. A feature of concealment is the increase in the proportion of descriptive stories, the enumeration of the furnishings; avoiding fixations on the faces of characters.

The generic results indicator have been developed. It is defined as the ratio of verbal to pre-verbal activity and may be quantified by calculating the ratio of the number of fixations on any stimulus before speaking to the number of words per story (F/W). The ratio of F/W to 3.0 corresponds to the freely true story; the ratio of F/W above 3.2 implies loss of speech production and is a specific marker of concealment.

Selected deception markers together with the previously identified "value-driven" eye-movement markers formed the basis of truth- and lie-detection technology for psychological testing. The special software have been developed by ErgoLab Ltd. for background measurement of an individual response rate in case of true answer. The software allows to set flexible criteria for identifying false responses according individual rate. Complex of eye movements shifts are compared to the rate during replies to each question in psychological testing process. Areas of increased value-driven attention and criteria for differentiation false and true answers can be used in applied research to reveal an true measure, even respondents try to hide it, for example in underlying motivation testing.

Acknowledgements

This research was sponsored by grant from the Russian Foundation for Basic Research (RFBR) according to the research project No. 15-06-10675a.

References

Cook A.E., Hacker D.J., Webb A.K., Osher D., Kristjansson S., Woltz D.J., & Kircher J.C. (2012). Lyin' Eyes: Ocular-motor Measures of Reading Reveal Deception. *Experimental Psychology,* 18(3), 301–313.

Ekman, P. (1992). *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage.* New York, London: Norton.

Ioffe S., Yesin S., & Afanasjev B. (2007). The Role of Subconscious Effects During the Treatment of Posttraumatic Stress Disorder with Alcohol Dependencies in Military Personnel. *Polygraph,* 36(3), 133–148.

Jarrett, C. (2012). The polygraph on trial. *The Psychologist,* Vol.2, No. 2, 104–105.

Naganuma, S., & Ochi, K. (2016). Examination of a presentation method for eye movement in the concealed information test. *International Journal of Psychology,* 51, 302.

Scott, G. (2009). *Emotion word processing: Evidence from electrophysiology, eye movements and decision making.* Submitted for the degree of PhD to the Higher Degree Committee of the Faculty of Information of Mathematical Sciences, University of Glasgow.

Serras Pereira M., Cozijn R., Postma E., Shahid S., & Swerts M. (2016). Comparing a Perceptual and an Automated Vision-Based Method for Lie Detection in Younger Children. *Frontiers in Psychology,* 12, 7:1936.

Vieira, K.M., Lane, S.M. (2013). How you lie affects what you remember. *Journal of Applied Research in Memory and Cognition,* 2, 173–178.

Vrij, A. (2008). *Detecting Lies and Deceit: Pitfalls and Opportunities,* Second Edition. Chichester, England: Wiley.

Walsh, D., & Bull, R. (2015). The association between interview skills, questioning and evidence disclosure strategies, and interview outcomes. *Psychology, Crime and Law,* 21, 661–680.

PARADOXICAL RELATIONSHIP BETWEEN SALIVARY OXYTOCIN CONCENTRATION AND AESTHETIC PREFERENCE FOR DRAWINGS: A PRELIMINARY STUDY

Hirokazu Doi and Kazuyuki Shinohara
Graduate School of Biomedical Sciences, Nagasaki University
<doi-h@nagsaki-u.ac.jp>

Abstract

It is well known that there are great individual differences in preference pattern for visual art forms. Many previous studies tried to elucidate the cause of such individual differences, but no study to date has clarified the relationship between neuro-endocrinological function and aesthetic preference pattern for visual art. The present study aimed to address this point by investigating the link between preference pattern for drawings and salivary concentration of oxytocin, a neuropeptide modulating human affiliative behavior. The results revealed a positive correlation between salivary oxytocin level and the strength of preference for drawings with still-life motif irrespective of style (abstract or realistic). Despite the previous findings that oxytocin modulates responses to human-related information, no such trend was observed in preference for portraits. These findings indicate that oxytocinergic neural function partly determines aesthetic preference pattern for drawings albeit in somewhat counterintuitive manner.

Drawing is one of the most popular art forms and as such many people enjoy viewing drawings/paintings in occasions such as art exhibitions. Drawings/paintings come in various styles like gothic, impressionism and cubism. In addition, painters can choose various motifs and paint them in their preferred style. Thus, combination of style and motif yields wide diversity in this art form.

It is well known that there are great individual differences in preference pattern for visual art forms (Chamorro-Premuzic, Burke, Hsu & Swami, 2010; Heinrichs & Cupchik, 1985). For example, some are strongly attracted to geometrical rendition of stringed instruments by cubists, while others find no aesthetic values in these pieces. Recent studies have shown that aesthetic experience induced by arts activates similar neural regions as erotic stimuli and high-calorie food, indicating the possibility that aesthetic experience is rooted in biological basis (Kirsch, Urgesi & Cross, 2016; Chatterjee, & Vartanian, 2014). Moreover, neural and behavioral responses to biologically-significant stimuli, related to survival and reproduction, have been shown to be under influences of endocrinological function (Scheele et al, 2013; Doi et al, 2017). On the basis of these, we hypothesized that individual differences in hormonal function underlie those in preference pattern for drawings.

The present study was designed to test this hypothesis by examining the relationship between trait-level of neuropeptide oxytocin and five types of drawings. Four categories of drawings were first defined by orthogonally combining motif (Portrait, Still Life) and style (Abstract, Realistic). We also presented drawings of Geometric category with no specific motif. Previous studies have indicated that oxytocin increases prosociality (Pfundmair et al, 2017; Shang, Wu & Su, 2017). Thus, we predicted that higher trait-level of oxytocin leads to increased preference to portrait painted in less abstract styles.

Method

Participants

22 females and 24 males (M=24.2, SD=7.05) with normal or corrected to normal visual acuity participated in the present study after giving written informed consent.

Procedure and Analysis

Five pieces were selected for each of the five types of drawings (Realistic-Portrait, Realistic-Sill Life, Abstract-Portrait, Abstract-Still Life, Geometrical). The drawings were chosen so that realistic panting and its abstract counterpart have similar composition. The abstract drawings included pieces from painters with various styles like abstract expressionism, fauvism and cubism.

Participants made subjective evaluation of preference for each of these 25 paintings using custom-made program. In each trial, a painting and vertical trackbar were presented on the display side-by-side. The upper most side of the trackbar was labelled "Like very much", while the lowermost side "Not at all". Participants answered their preference by moving the trackbar to the location closest to their feeling.

The participants' saliva samples were collected into 1.5 ml tube by passive drool. We collected the samples during 12:00–14:00 in order to mitigate the influences of diurnal fluctuation in hormone level. The participants were refrained from tooth blush and eating 1hr before saliva collection and rinsed their mouth just before saliva donation. In the analysis, the averaged evaluations for drawings in each category were calculated. We assayed the salivary oxytocin level by ELISA using a commercial kit.

Results

The averaged evaluation for each type of drawing was regressed against salivary oxytocin level. Because of explanatory nature of this analysis, we did not control for the family wise error rate. The correlational coefficients are summarized in Table 1. As can be seen in this table, preference evaluations for drawings of Still Life showed significantly positive correlation with salivary oxytocin concentration irrespective of style, $ps < .05$.

Discussion

The present study investigated the association between preference pattern for five types of drawings and trait-level salivary oxytocin concentration. The analysis revealed that those with higher level of salivary oxytocin show stronger preference to drawings of still life irrespective of painting style, i.e. realistic or abstract. This partially supports our hypothesis that individual differences in endocrinological function influence the pattern of aesthetic preference. At the same time, considering the characterization of oxytocin, the present preliminary findings seem a bit counterintuitive. As such, further, more systematic, investigation is required to elucidate the biological basis of individual differences of visual art preference.

Table 1. Averaged evaluations and correlational coefficients (R_s) between salivary oxytocin concentration and evaluations for each category of drawings. In the parenthesis are the standard deviations.

Category	Evaluation	R
Realistic-Still Life	12.3 (2.1)	0.296*
Abstract-Still Life	8.8 (2.7)	0.298*
Realistic-Portrait	11.1 (3.0)	0.087
Abstract-Portrait	7.1 (2.5)	-0.016
Geometric	9.7 (2.5)	0.186

*$p < .05$

References

Chamorro-Premuzic, T., Burke, C., Hsu, A., & Swami, V. (2010). Personality Predictors of Artistic Preferences as a Function of the Emotional Valence and Perceived Complexity of Paintings. *Psychology of Aesthetics, Creativity, and the Arts,* 4(4), 196–204.

Heinrichs, R.W., & Cupchik, G.C. (1985). Individual differences as predictors of preference in visual art. *Journal of Personality,* 53(3), 502–515.

Kirsch, L.P., Urgesi, C., & Cross, E.S. (2016). Shaping and reshaping the aesthetic brain: Emerging perspectives on the neurobiology of embodied aesthetics. *Neuroscience and Biobehavioral Reviews,* 62, 56–68.

Chatterjee, A., & Vartanian, O. (2014). Neuroaesthetics. *Trends in Cognitive Sciences,* 18 (7), 370–375.

Scheele, D., Wille, A., Kendrick, K.M. , Maier, W., & Hurlemann, R. (2013). Oxytocin enhances brain reward system responses in men viewing the face of their female partner. *Proceedings of the National Academy of Sciences of the United States of America,* 110(50), 20308–20313.

Doi, H., Morikawa, M., Inadomi, N., Aikawa, Uetani, M., & Shinohara, K. (2017). Neural correlates of babyish adult face processing in men. *Neuropsychologia,* 9–17.

Pfundmair, M., Zwarg, C., Paulus, M., & Rimpel, A. (2017). Oxytocin promotes attention to social cues regardless of group membership. *Hormones and Behavior,* 90, 136–140.

Shang, S., Wu, N., & Su, Y. (2017). How oxytocin receptor (OXTR) single nucleotide polymorphisms act on prosociality: The mediation role of moral evaluation. *Frontiers in Psychology,* 8(MAR), 396.