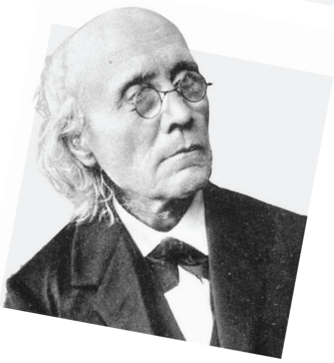
34th Annual Meeting of the International Society for Psychophysics

20th – 24th August 2018

CONFERENCE PROCEEDINGS



FECHNER DAY 2018
LÜNEBURG

Proceedings of the 34th Annual Meeting of the International Society for Psychophysics

Fechner Day 2018

Lüneburg, Germany
20–24 August 2018

EDITORS

Friedrich Müller, Lara Ludwigs, and Malizia Kupper
Leuphana University Lüneburg

REFERENCE

Müller, Friedrich; Ludwigs, Lara; and Kupper, Malizia (Eds.). *Fechner Day 2018: Proceedings of the 34th Annual Meeting of the International Society for Psychophysics*. International Society for Psychophysics, Lüneburg, Germany, 2018.

DISTRIBUTION

International Society for Psychophysics, c/o Prof. Dr. Friedrich Müller
Institute of Experimental Industrial Psychology - LueneLab,
Leuphana University, Wilschenbrucher Weg 84a, D-21335 Lüneburg, Germany
Email: f.mueller@leuphana.de

ACKNOWLEDGEMENTS

Fechner Day 2018 is generously sponsored by

German Research Foundation

Ministry for Science and Culture of Lower Saxony

Leuphana University Lüneburg

Faculty of Business and Economics, Leuphana

Printed by

VARIOPAPER GmbH, Wichernstraße 34, D-21335 Lüneburg, Germany

International Society for Psychophysics

Founded in Cassis, France, 22 October 1985.

Executive Committee

President	John Monahan (Central Michigan University, Mount Pleasant, USA)
Vice-President	Friedrich Müller (Leuphana University Lüneburg, Lüneburg, Germany)
Treasurer	Wolfgang Ellermeier (Technische Universität Darmstadt, Darmstadt, Germany)
Secretary	Timothy Hubbard (Arizona State University, Tempe, AZ, USA)

Organising Committee of Fechner Day 2018

Chair	Friedrich Müller (Leuphana University Lüneburg, Lüneburg, Germany)
	Lara Ludwigs (Leuphana University Lüneburg, Lüneburg, Germany)
	Malizia Kupper (Leuphana University Lüneburg, Lüneburg, Germany)

Programme Committee of Fechner Day 2018

Wolfgang Ellermeier (Technische Universität Darmstadt, Darmstadt, Germany)

Leah Fostick (Ariel University, Ariel, Israel)

Friedrich Müller (Leuphana University Lüneburg, Lüneburg, Germany)

Bruce A. Schneider (University of Toronto Mississauga, Ontario, Canada)

Jordan Schoenherr (Carleton University, Ottawa, ON, Canada)

Previous Annual Meetings

YEAR	VENUE	ORGANISER(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, NC, 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. RiberoFilho

YEAR	VENUE	ORGANISER(S)
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
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
2017	Fukuoka, Japan	Y. Nakajima, K. Ueda, and G. B. Remijn

Contents

Welcome to Fechner Day 2018

The first Fechner Day I was able to participate in was the meeting in 1987 organised by Mimi and Bob Teghtsoonian in the forests of New Hampshire. We, at the time rather young scientists, were not only enthusiastic about the overwhelming Indian Summer. We were especially attracted by the friendly reception and acceptance by the established scientists and were impressed by the informal but very intensive exchange of scientific ideas. An atmosphere that has been preserved to these days.

When, inspired by these experiences, we decided to invite to Fechner Day 1990 in Würzburg, the oldest psychological institute after Leipzig, we started very early in advance to get permissions for colleagues from the Eastern Bloc countries to participate. Among other requirements, I had to go to the Foreigners' Registration Office to sign a commitment letter to provide a living for colleagues who might decide to not return to their home country and to stay in the West.

When the meeting finally took place, the Cold War was over. The wall between the two parts of Germany had fallen and the colleagues could travel unhindered to Fechner Day. Never before have I been so happy to have invested effort in, in retrospect, superfluous activities. Thanks to the support of the German Research Foundation (DFG) and the Bavarian Ministry of Research, Fechner Day 1990 was held in a very optimistic mood.

The political developments at that time promoted the emergence of the university hosting this year´s Fechner Day. During the Cold War, Lüneburg was situated close to the Iron Curtain and surrounded by several barracks. The military was the most important economic influence for the region. With perestroika, the soldiers left. As a compensation, Lüneburg got a university and the already existing university of applied sciences was considerably expanded. Both were merged in 2005 and one of Lüneburg's largest barracks was converted into a spacious and now very friendly university campus. Due to this conversion, the increasing number of students also stimulated and influenced the social and cultural life in the town. There is hardly any other historic town in Germany as lively and as full of ideas as Lüneburg. We believe that this atmosphere will inspire the Fechner Day 2018 as well.

Scientific impulses are once again guaranteed by the contributions of the members. This year's contributions cover almost the entire field of research in our discipline. Some contributions are committed to the traditional task of psychophysics of developing new measurement systems and experimental methods for psychological research, while others look into the theoretical founding of existing ones. Studies from basic and applied research in various fields of perception and general psychology prove that psychophysical methods are invaluable for science and were not invented “just to pester first-year students”.

Leading scientists, involved in the development and implementation of ultra-modern technologies at the interface between brain and technical devices, will point out current challenges for psychophysical research and the need for suitable psychophysical measures and measuring procedures. By doing so, they will demonstrate once again that the oldest psychological discipline will probably not be on the red list of dying out sciences for a long time to come. In Walter Boland's observations, psychophysical processes can be discovered in the communication of plants.

Fechner Day 2018 is generously supported by the DFG, the Ministry of Science and Culture of Lower Saxony, the Leuphana University and the Faculty of Business and Economics. We are very grateful for their support which obliges us to do our best to provide the ground for a successful conference.

Preparing and holding a conference is teamwork. I am happy that interested and committed students joined the organising team and accompany the conference.

My very special thanks go to Lara Ludwigs and Malizia Kupper who made Fechner Day their own concern. Without their highly professional commitment, Fechner Day 2018 would and could not be held at the Leuphana University in Lüneburg.

We hope that we will succeed in creating the framework for an interesting conference. And, we hope that we are able to create the friendly, socially relaxed, stimulating and supporting atmosphere when presenting facts and discussing controversial thoughts, as it is characteristic of all previous Fechner Days.

Lüneburg, 20th August 2018
Friedrich Müller

INVITED LECTURES

MOTHER NATURE'S MAGIC BULLETS: ENHANCING THE TERPENOID ARSENAL BY ENVIRONMENTAL CUES

Wilhelm Boland
Max Planck Institute for Chemical Ecology, Jena, Germany

IDEAL OBSERVER MODELS IN COGNITIVE SCIENCE: PROGRESS OR FAD?

Frank Jäkel
Technische Universität Darmstadt
jaekel@psychologie.tu-darmstadt.de

Bayesian models of cognition and behavior compare human performance to the performance of an ideal decision maker. Such models are particularly promising when they are used in reverse-engineering explanations: explanations that descend from Marr's computational level of analysis to the algorithmic and the implementational level. Unfortunately, it remains unclear exactly how Bayesian models constrain and influence these lower levels of analysis. In several examples ranging from signal detection tasks to inductive reasoning I sketch and compare how Bayesian models are used in psychophysics and cognitive science.

SYMPOSIUM: BIOELECTRONICS

Invited speakers.

ADVANCES WITH COCHLEAR IMPLANTS

Bernhard Laback
Acoustics Research Institute, Austrian Academy of Sciences

The cochlear implant (CI) is considered as the most successful sensory prosthesis. While originally developed to restore some basic hearing sensation in the deaf, current clinical CI systems allow recipients to achieve speech recognition rates of about 80% for sentences and of about 50% for monosyllabic words. In more realistic listening situations involving interfering sounds, however, recognition rates decline dramatically, causing a substantial performance deficit compared to normal-hearing listeners. Furthermore, spatial hearing and music perception abilities are largely impaired with current clinical CI devices, even when implanted on both ears. This talk will give an overview of the basic approach, current challenges, and promising developments with CIs. The focus will be given to potential reasons for performance deficits and strategies to overcome them from the perspective of psychophysics.

RETINAL IMPLANTS: FROM FIRST CONCEPTS TO FIRST PATIENTS AND BEYOND

Peter Walter, Tibor Lohmann, Kim Schaffrath, and Sandra Johnen
Department of Ophthalmology, RWTH Aachen University

The concept of a retinal implant was introduced more than 40 years ago. In the meantime, a lot of technological and surgical advances took place. The idea of a retinal implant is to bridge the input of photoreceptors in patients where these cells are dead as is the case in Retinitis pigmentosa or Macular Degeneration using electrostimulation with implantable flexible and biocompatible multielectrode arrays. The electrode arrays can be placed in the suprachoroidal space, in the subretinal space or onto the retinal surface targeting different retinal neurons. The specific pros and cons of the different approaches will be discussed. When implanted into blind patients about 50% reported a benefit in terms of better mobility and the perception of strong contrast changes helping these patients to identify and locate obstacles or target objects however the perception of form or movement is difficult. In the last years we learned a lot of the retinal networks in degeneration and how these changes may interfere with "simple" electrostimulation strategies. A few examples will be given from mouse models of retinal degeneration and what we may learn from these experiments. Finally, an outlook of the technology in comparison to other forms of treatment for "untreatable blindness" such as gene therapy or stem cell technology will be given.

SOMATOSENSORY FEEDBACK PROSTHESES

Caroline Dietrich and Thomas Weiss
Department of Clinical Psychology, Friedrich Schiller University Jena
caroline.dietrich@uni-jena.de

After amputation of a limb, sensations of touch and movement are inherently lost. Despite the physical loss, the majority of amputees continues to perceive the limb (phantom limb), often as painful (phantom limb pain). This talk reviews the possibilities of current limb prostheses to replace motor and especially somatosensory functions of a lost limb. In detail, the talk focuses on the role of somatosensory feedback prostheses in affecting phantom limb pain and improving prosthesis control.

SYMPOSIUM:
AUDITORY TEMPORAL PROCESSING

Organised by Leah Fostick and Harvey Babkoff.

AUDITORY TEMPORAL PROCESSING: EVIDENCE FOR MULTIPLE PHENOMENA

Leah Fostick
Department of Communication Disorder, Ariel University, Ariel, Israel
Harvey Babkoff
Department of Psychology, Bar-Ilan University, Ramat-Gan, Israel
Leah.fostick@ariel.ac.il

Abstract

The temporal order judgment (TOJ) paradigm is used to measure auditory temporal processing by asking the participant to report the order of two sounds. In the present study, we tested whether temporal processing is dependent upon the characteristics of the stimuli to be judged by comparing the TOJ threshold distributions from different two-sound paradigm studies that manipulated a variety of auditory stimulus parameters. The 52 TOJ experiments provided a large variety of thresholds, ranging from 0 msec (accuracy rate > 75% even when there is no temporal interval separating the offset of the first from the onset of the second stimulus (ISI)) to ISI > 200 msec (accuracy rate < 75% at the longest ISI value). These threshold ranges were divided into three categories: (1) ISI threshold < 5 msec (the shortest ISI tested in the experiments), representing correct judgment of order without any temporal separation between the sounds; (2) ISI thresholds throughout the range 5 msec < ISI threshold < 200 msec; (3) ISI threshold > 200 msec, representing participants with accuracy rate < 75% even at the longest ISI value. The results indicated that the stimulus parameters in two-sound TOJ paradigms have a strong influence on TOJ threshold distributions. These data support the hypothesis of multiple mechanisms that can produce correct judgments of temporal order and do not support the hypothesis that posits a single central temporal order mechanism. In addition, we also conclude that the TOJ paradigm most capable of producing correct temporal order judgments even when ISI = 0 msec, is the paradigm of two tones that differ in frequency. Furthermore, the success of correctly judging temporal order even when ISI=0 msec, depends upon the pitch of the tones, whether they are low, medium or high.

Auditory temporal processing refers to the ability of the auditory nervous system to temporally process rapidly presented sounds. The implicit assumption for using auditory TOJ tasks when comparing temporal processing across a variety of sub-populations is that the TOJ paradigm measures temporal processing, similar to other auditory tasks that also manipulate the temporal parameter, e.g., duration discrimination, backward masking, and gap detection. However, each one of these auditory tasks might by influenced by a variety of auditory processing mechanisms, that are not temporal, such as hearing sensitivity and frequency discrimination. Several earlier researchers suggested that the same central mechanism is responsible for temporal order judgments, regardless of the stimulus modality or method of presentation. The evidence for this hypothesis is mixed and has been challenged by recent studies. Hirsh and colleagues (Hirsh, 1959; Hirsh and Sherrick, 1961) used repeated two-sound sequences to test the effects of different types of stimuli on temporal order judgments. They reported that no matter what types of stimuli they used (tones, click, visual or tactile stimuli), stimulus-onset asynchrony (SOA) thresholds were approximately the same (i.e., 17 msec). However, Warren et al. (e.g., Warren and Ackroff, 1976; Warren and Bashford, 1993; Warren and Byrnes, 1975; Warren and Obusek, 1972; Warren, 1974a,b) found that order judgments changed in accordance with the type of stimuli utilized in the sequence. More recently, researchers also showed differences in response patterns when the TOJ paradigm used two different frequencies than when the same frequency tones were presented asynchronuously to the two ears (Fink, Churan, & Wittmann, 2005, 2006; Fink, Ulbrich, Churan, & Wittmann, 2006; Fostick and Babkoff, 2013, 2017; Szymaszek, Sereda, Pöppel, & Szelag, 2009).

The aim of the present study was to test whether auditory temporal order judgment is independent of the characteristics of the stimuli to be judged, by comparing TOJ thresholds to a variety of different two-sound paradigms, with different stimulus parameters. If the TOJ paradigms with different parameters yield different response patterns, this would be strong evidence for the involvement of more than one perceptual mechanism for the judgment of temporal order. If the same or similar threshold distributions are generated by a variety of different stimulus parameters, then one might argue more cogently for the hypothesis that temporal order judgments are mainly performed by the same central mechanism. To test the two hypotheses, we compared the TOJ threshold distributions when the two auditory stimuli were either: 1) of two different frequencies (low and high tones); or 2) a pure tone and wide-band noise; or 3) both tones were the same frequency but differed in intensity (low and high sensation level); 4) both tones were the same frequency and intensity, but differed by duration (short and long tones);or 5) both tones were the same frequency, the same intensity and the same duration, but were presented dichotically and differed by which ear (right or left) was stimulated first and which ear was stimulated second (location).

Method

Participants

This study is based on 1,286 participants (58% females) who participated in 52 different TOJ experiments. All participants were University undergraduates, age 20 – 35 years who were screened for normal hearing.

Stimuli, tasks, and procedure

All experiments were approved by the University Ethics Committee. The participants provided signed informed consent before beginning the experimental procedure. In all the experiments, participants were presented with two sounds and were asked to report which sound was first and which, second. The sounds differed in one of the following parameters: (1) frequency (participants' response: high-low or low-high); (2) location of the ear to which the first sound was presented (participants' response: right-left or left-right); (3) spectrum of the sound [being either pure tone or Gaussian noise (participants' response: tone-noise or noise-tone)]; (4) intensity (participants' response: soft-loud or loud-soft); and (5) duration (participants' response: short-long or long-short). Participants were trained to recognize the stimuli and perform order judgments before starting the experiment (see Fostick & Babkoff, 2013 for detailed description of the training procedure). The experiments were performed either by the constant stimulus method or by the adaptive procedure (see Fostick and 2017 for detailed description of these methods). The experimental method, mode of presentation, frequency, duration, and intensity TOJ parameters are shown in Table 1.

Table 1. Characteristics of the experiments for each parameter

Parameter	k	n	Method	Presentation	Frequency (Hz)	Stimulus length (msec)	Intensity
Frequency	28	693	Constant stimuli, adaptive procedure	Synchronous, asynchronous	200-400, 250-500, 300-600, 400-800, 500-1,000, 510-550, 510-650, 600-1,200, 750-1,500, 1,000-1,100, 1,000-1,500, 1,000-1,800, 1,000-2,000, 1,000-3,500 2,500-5,000, 2,750-5,500, 3,000-6,000	15, 75	40dBSL, 60dBSPL
Location	20	465	Constant stimuli, adaptive procedure	Asynchronous	300, 600, 1000, 1500	10, 15, 20, 30, 40	40dBSL, 60dBSPL
Spectrum	1	34	Adaptive procedure	Synchronous	1,000-GN	15	40dBSL
Duration	2	59	Adaptive procedure	Synchronous	1000	10-30, 15-45	40dBSL
Intensity	1	35	Adaptive procedure	Synchronous	1000	15	25dBSL-40dBSL

Since stimulus duration was one of the experimental manipulations, TOJ threshold was measured as the inter-stimulus-interval (ISI), the amount of time that separates the offset of the first stimulus from the onset of the second stimulus (and not stimulus-onset asynchrony (SOA) that is often reported as TOJ threshold in some studies). Participants' threshold was calculated as the ISI for 75% correct responses.

Results and Discussion

TOJ is used to measure auditory temporal processing by asking the participant to report the order of two sounds. In the present study, we tested whether temporal processing is dependent upon the characteristics of the stimuli to be judged. The 52 TOJ experiments provided a large variety of thresholds, ranging from ISI thresholds of 0 msec (accuracy rate > 75% even when there is no time separating the offset of the first from the onset of the second stimulus) to ISI > 200 msec (accuracy rate < 75% at the longest ISI value). These thresholds were divided into three categories: (1) ISI threshold < 5 msec (the shortest ISI tested in the experiments), representing performance without any temporal separation between the sounds; (2) thresholds throughout the range of ISIs (5 msec < ISI threshold < 200 msec); (3) ISI threshold > 200 msec, representing participants with accuracy rate < 75% even at the longest ISI value. The association between TOJ parameters and threshold categories was tested using Cramer's V for measuring association between two nominal variables (Cramér, 1946). Due to the large number of participants in some of the analyses, the effect size was used to interpret statistically significant results (Cohen, 1988).

1. Between-parameters analysis

Participants judged the order of two tones that differed either in their frequency, location (order of ears stimulated), spectrum width, duration, or intensity. Figure 1 shows the distribution of TOJ threshold categories for these parameters. Cramer's V showed significant associations between parameter type and threshold category (φ_c = .425, $p = .000$). These results show that although all of the experiments measured TOJ as the dependent variable, the different stimulus parameters yielded different threshold distributions. The threshold distribution that appears most different from the rest is the one generated by Frequency (Fig. 1).

Frequency and Location were previously found to have different threshold distributions (Fostick, & Babkoff, 2013, 2017). In the present study a much larger number of participants showed the same tendency with a very large effect size (φ_c = .595, $p = .000$). The other parameters also yielded threshold distributions that differed from those yielded when the frequencies differed or when the same stimuli were presented asynchronously to the two ears (location) (see Table 2). However, although all analyses were statistically significant, the effect sizes were different. There were medium effect sizes when Frequency was compared with Spectrum and Intensity, while there was a large effect size when Frequency was compared with Duration. Only small to medium effect sizes were found when Location was compared with Spectrum and Duration, while a large effect size was found when Location was compared with Intensity. These results mean that, in general, Frequency provides very different threshold distributions than Location and Duration, and that Location provides very different threshold distributions than Frequency and Intensity. Stimulus manipulations involving duration, intensity and spectrum do not seem to yield significantly different TOJ threshold distributions (much lower effect sizes). However, this might be due to the relatively small number of participants in those experiments and, consequently, the larger variance.

2. Within-parameters analysis

Since major differences were found when we compared the TOJ threshold distributions yielded when the paradigm consists of two pure tones of different frequencies (frequency), to the paradigm that consists of the same tones presented asynchronously to the two ears (location), we also compared the effects of the frequency location along the spectrum and the temporal length of the tones on the two paradigms: Frequency and Location. All Cramer's V associations were statistically significant. However, there were differences in effect sizes.

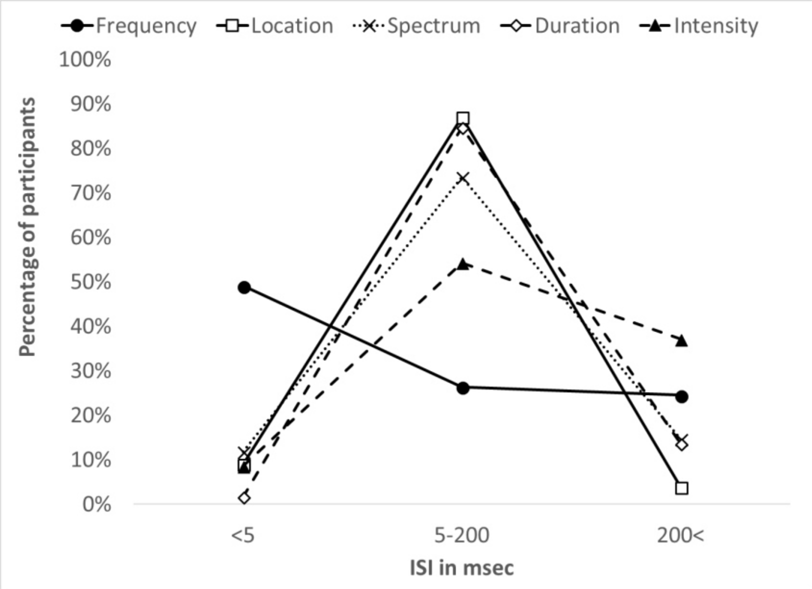


Figure 1. Distribution of TOJ threshold categories for TOJ based on differences in Frequency, Location, Spectrum, Duration, and Intensity

Table 2. Cramer's V results for testing the association between TOJ parameters of Frequency and Location with Spectrum, Duration, and Intensity.

	Spectrum	Duration	Intensity
Frequency	.222***	.345***	.178***
Location	.134*	.160**	.353***

* $p < .05$; ** $p < .01$; *** $p < .001$

For the Frequency paradigm, there were large effect sizes, indicating a strong relationship between TOJ threshold distributions and whether the tone was low, medium or high in pitch ($\varphi_c = .302$, $p = .000$). There were only very small to medium effect sizes when we compared longer and shorter tone lengths ($\varphi_c = .172$, $p = .000$). The opposite picture emerged for the Location paradigm. It made very little difference whether the tones were low or medium in pitch ($\varphi_c = .122$, $p = .032$); whereas the temporal length of the tones produced large effect sizes ($\varphi_c = .332$, $p = .000$).

Conclusions

The results of all of the comparisons indicate that the stimulus parameters in two-sound TOJ paradigms have a strong influence on the TOJ threshold distributions. Consequently, we find that the data of a large number of experiments do not support the hypothesis of a single temporal order mechanism for all types of auditory stimuli. Rather, it would seem that there are multiple mechanisms that can produce correct judgments of temporal order. In addition, we also conclude that the TOJ paradigm most capable of producing correct temporal order judgments even when ISI = 0 Msec, is the paradigm of two tones that differ in frequency. Furthermore, the success of correctly judging temporal order even when ISI=0 Msec, depends upon the pitch of the tones, whether they are low, medium or high.

References

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cramér, H. (1946). Mathematical Methods of Statistics. Princeton: Princeton University Press.

Fink, M., Churan, J., Wittmann, M. (2005). Assessment of auditory temporal-order thresholds – A comparison of different measurement procedures and the influences of age and gender. *Restorative Neurology and Neuroscience, 23*, 281–296

Fink, M., Churan, J., Wittmann, M. (2006). Temporal processing and context dependency of phoneme discrimination in patients with aphasia. *Brain and Language, 98*, 1–11.

Fink, M., Ulbrich, P., Churan, J., Wittmann, M. (2006). Stimulus-dependent processing of temporal order. *Behavioural Processes, 71*, 344–352.

Fostick, L., Babkoff, H. (2013). Different Response Patterns between Auditory Spectral and Spatial Temporal Order. *Experimental Psychology*, *60*(6) 432–443.

Fostick, L., Babkoff, H. (2017). Auditory spectral versus spatial temporal order judgment: Threshold distribution analysis. *Journal of Experimental Psychology: Human Perception and Performance, 43*(5), 1002–1012.

Hirsh, I. J. (1959). Auditory perception of temporal order. *The Journal of the Acoustical Society of America*, *31*(6), 759-767.

Hirsh, I. J., & Sherrick Jr, C. E. (1961). Perceived order in different sense modalities. *Journal of experimental psychology*, *62*(5), 423.

Szymaszek, A., Sereda, M., Pöppel, E., Szelag, E. (2009). Individual differences in the perception of temporal order: The effect of age and cognition. *Cognitive Neuropsychology, 26*(2), 135–147.

Warren, R. M. (1974a). Auditory temporal discrimination by trained listeners. *Cognitive Psychology*, *6*(2), 237-256.

Warren, R. M. (1974b). Auditory pattern recognition by untrained listeners. *Perception & Psychophysics*, *15*(3), 495-500.

Warren, R. M., & Bashford, J. A. (1993). When acoustic sequences are not perceptual sequences: The global perception of auditory patterns. *Perception & psychophysics*, *54*(1), 121-126.

Warren, R. M., & Byrnes, D. L. (1975). Temporal discrimination of recycled tonal sequences: Pattern matching and naming of order by untrained listeners. *Perception & Psychophysics*, *18*(4), 273-280.

Warren, R. M., & Obusek, C. J. (1972). Identification of temporal order within auditory sequences. *Perception & Psychophysics*, *12*(1), 86-90.

Warren, R. M., Ackroff, J. M. (1976). Two types of auditory sequence perception. *Perception and Psychophysics, 20*, 387–399.

TIME WINDOWS IN SENSORY PROCESSING AS NEURAL BASIS FOR COGNITIVE FUNCTIONS

Ernst Pöppel [1, 2], Taoxi Yang [1], Bin Zhou [3], and Yan Bao [1, 2]
[1]*Institute of Medical Psychology, Ludwig-Maximilian-University Munich, Germany;*
[2]*School of Psychological and Cognitive Sciences, Peking University, China;*
[3]*Institute of Psychology, Chinese Academy of Sciences, China*
ernst.poeppel@med.uni-muenchen.de, taoxi.yang@campus.lmu.de,
zhoub@psych.ac.cn, baoyan@pku.edu.cn

One basic question in cognitive science is (or should be) whether information processing in sensory systems is continuous or discrete; it appears that the distinction between the two processing modes has remained largely an "unasked question". There are several reasons to consider discrete temporal processing using time windows (which are implemented in the neural systems on a pre-semantic level) as processing mode within and across sensory modalities: 1) Time windows would overcome with respect to event detection a neuro-anatomical problem; because of the axonal divergence in the projection, the arrival time of afferent information at central areas like the sensory cortices is temporally ill-defined. Within time windows (which on the basis of our reasoning are implemented by the period of a "relaxation oscillation") afferent information is treated as co-temporal, i.e. the concept of continuity of time (like in classical physics) does not apply within time windows. 2) A pre-semantically implemented time window being created by oscillatory processes would allow a better understanding for sequential processing stages; one essential problem to be solved in cognition is when an event has come to an end to allow the beginning of the next processing stage. A pre-semantic mechanism as reflected in time windows would liberate the system from an ill-understood top-down control. 3) The transduction time in sensory systems is different being for instance much shorter in the auditory system compared to the visual system; for inter-sensory integration time windows of finite duration appear to be useful as the central availability of information is temporally distributed. 4) The physical distance of objects to be perceived in different sensory spaces results in temporal uncertainty between the modalities; at the "horizon of simultaneity" of some 10 meter distance the transmission time of sound corresponds to the transduction time in the retina; up to this horizon auditory stimuli are represented earlier at the cortical level; beyond this horizon visual stimuli; this creates another challenge for intersensory integration. It is concluded that these physical, biophysical, anatomical, physiological and cognitive challenges can be overcome by discrete time sampling. Empirical evidence with different experimental paradigms supports the notion for time windows with an operating range from 30 to 40 ms, or more conservatively speaking from 20 to 60 ms, a prime example being measurements on temporal order threshold in different sensory modalities. Other evidence comes from multimodal distributions of simple or choice reaction to auditory and visual stimuli, latency of pursuit or saccadic eye movements, temporal tolerance in stereopsis, temporal tolerance in sensorimotor synchronization, auditory mid-latency evoked responses, or single cell activities in sensory pathways. The numerical similarity of the operating range with these different experimental paradigms supports the theoretical concept of discrete time sampling, in particular the existence of time windows being implemented by relaxation oscillations in neural systems.

Reference

Bao, Y., Pöppel, E., Wang, L., Lin, X., Yang, T., Avram, M., Blautzik, J., Paolini, M., Silveira, S., Vedder, A., Zaytseva, Y., & Zhou, B (2015). Synchronization as a biological, psychological and social mechanism to create common time: A theoretical frame and a single case study. *PsyCh Journal* 4: 243–254. DOI: 10.1002/pchj.119

MODULATION OF TEMPORAL ORDER JUDGMENT IN AUDITION

Yan Bao, Zhilin Zhang, and Xiaoxiong Lin
School of Psychological and Cognitive Sciences, Peking University, Beijing 100871, China
baoyan@pku.edu.cn, zhangzhilin@pku.edu.cn, lxx.aesthetics@gmail.com

The link between temporal perception and language processing is traditionally considered as one directional, i.e., the ability to perceive the temporal order of successive acoustic events provides a fundamental basis for understanding speech with its rapidly changing spectro-temporal patterns. However, recent studies measuring temporal order threshold of two tones with different frequencies suggest that temporal processing is also modulated by the long-term exposure to a native language environment (Bao et al., 2013) and even by mastering a distinctively different secondary language (Bao et al., 2014). Specifically, native Chinese speakers show significantly lower temporal order thresholds in discriminating the order of two close frequency tones (600Hz and 1200Hz), while native Polish speakers show the opposite, i.e., lower temporal order thresholds in discriminating two distant frequency tones (400 Hz and 3000 Hz). Compared to native Chinese without non-tonal language experience, Chinese subjects with English proficiency show lower temporal order thresholds in discriminating both close frequency and distant frequency tones. These observations suggest language experience as a modulating factor for auditory temporal order perception, and they suggest a bi-directional relationship between temporal and language processing. To explain the different performance associated with the two types of tone pairs, two processing modes or strategies have been proposed for decoding the sequence of two successive events (Bao et al., 2013). One mode is analytic, which refers to experiencing temporal order by identifying distinct acoustic events; the other mode is holistic, which refers to extracting the temporal order on the basis of a global perceptual pattern generated by integrating the two successive tones into a unitary percept with either an upward or downward acoustic patterning. The latter presumably relies on auditory neurons responding selectively to the direction of a spectral change. To further test these two processing modes, we conducted a psychophysical experiment on temporal order judgment using a sensory adaptation paradigm, i.e., presenting a series of upward or downward frequency change adaptors before the tone pairs appeared. The results showed a significant adaptation effect as indicated by the difference of the points of subjective simultaneity (PSS) between the downward and upward adaptor conditions, however, mainly for temporal order judgment of close frequency tones. These results confirm a holistic processing mode for close frequency tones which perceptually can easily be integrated for our sample of Chinese subjects. Whether the holistic processing mode can also be demonstrated using the sensory adaptation paradigm in subjects who can easily integrate distant frequency tones remains an open question.

References

Bao Y, Szymaszek A, Wang X, Oron A, Pöppel E, & Szelag E. (2013). Temporal Order Perception of Auditory Stimuli is Selectively Modified by Tonal and Non-tonal Language Environments. *Cognition*, 129: 579-585.

Bao Y, Fang Y, Yang T, Wang L, Szymaszek A, & Szelag E. (2014) Auditory perception of temporal order: A comparison between tonal language speakers with and without non-tonal language experience. *Acta Neurobiologiae Experimentalis*, 74, 98-103.

LOCALIZING SOUNDS IN TIME: A NEW TIMING TASK TO PREVENT CHRONOMETRIC COUNTING?

Martin Riemer[1,2], Paula Vieweg[1,3], and Thomas Wolbers[1]
[1]*Aging & Cognition Research Group, German Center for Neurodegenerative Diseases (DZNE), Magdeburg, Germany*
[2]*Medical Faculty (FME), Otto-von-Guericke University, Magdeburg, Germany*
[3]*Department of Experimental Psychology and Methods, University of Leipzig, Germany*
martin.riemer@dzne.de

Chronometric counting is a prevalent issue in psychophysical studies on time perception. As researchers are often interested in pure timing abilities, different methods have been suggested to prevent participants from chronometric counting. All of these methods possess both pros and cons and have successfully been applied in various studies. Nevertheless, an ideal situation would consist in a time perception task which inherently does not trigger a counting strategy.

We presented our participants with a short auditory event within a variable reference duration and asked them to judge whether the event occurred in the first or the second half of the reference duration. We hypothesized that this experimental design reduces the utilization of counting strategies, because (i) the instruction explicitly requires to ‘localize an event in time’, rather than to ‘compare two separate intervals’, and (ii) the direct transition from the first to the second half of the reference duration interferes with the strategy to start counting from zero for both interval halves.

The comparison with a conventional discrimination task, in which participants indicate which of two intervals is longer in duration, reveals that event localization in time is significantly less affected by explicitly instructed counting, relative to ‘no-counting’ conditions.

Event localization in time represents a promising approach to study time perception without the confound of chronometric counting.

TEMPORAL PROCESS: WHAT IT CAN TELL US

Eliane Schochat
Faculdade de Medicina FMUSP, Universidade de Sao Paulo, Sao Paulo, SP, BR
eschocha@usp.br

Auditory temporal processing refers to the ability of the central auditory nervous system (CANS) to encode and detect subtle changes in acoustic signals, and normal temporal processing is necessary for perception of these acoustic changes. This processing has been further identified as being critical for normal auditory development (Moore, 2003).

Auditory temporal processing can be didactically divided in four categories: temporal ordering, temporal resolution, temporal integration, and temporal masking (Shinn, 2003). Nowadays, there are available and well-documented clinical measures to assess temporal ordering and resolution, but not temporal integration and masking (Filippini, Schochat, 2014).

In the clinic, temporal processing can be evaluated behaviorally through measures such as the gaps-in-noise (GIN) test (Musiek et al, 2005), pitch pattern test (PPT) and duration pattern test (DPT). The GIN test was developed by Musiek in 2005 to characterize the ability of the CANS to resolve very brief acoustic changes. The PPT and DPT were also developed by Musiek (1994) and is devoted to analyze one´s ability to temporal order the tones.

Temporal Processing can also be defined by the rate at which we can process auditory information. A person must be able to process auditory information at a rapid pace in order to develop appropriate listening and language skills. Audiologists have recognized this in people with sensorineural hearing loss for a long time and have referred to this concept as the "temporal window". We know that if a person's "temporal window" is too large, that is, the time period required to process sound is too long; it becomes more difficult for them to understand speech. Any brief change in the speech signal then becomes difficult to perceive and the communication is distorted.

Since the 1990s, research has supported the hypothesis, initially proposed by Tallal & Piercy (1973), that language disorders are related to a deficit in auditory temporal processing. According to Habib (2000), difficulties are observed in the processing of the temporal characteristics of different types of sensory stimuli, including auditory, visual and sensory-motor stimuli, when the stimuli are presented in rapid succession. More specifically, difficulty involving auditory temporal processing is expressed as a limited capacity to process ''short acoustical elements'', such as consonants, that comprise the rapid transition of formants. Limitations in this capacity can lead to difficulties, such as associating letters with their specific sounds, which can potentially result in dyslexia.

As it is now well-known speech perception depends on temporal features of the speech signal. Rosen (1992) presented a framework for these temporal features that segregated the speech signal into three frequency ranges: the low-frequency speech envelope (2-50 Hz), periodicity cues (50-500 Hz), and the temporal fine structure (600-10,000 Hz) (Rosen, 1992). The low-frequency envelope is dominated by the syllable rate of speech and is sufficient for speech perception in quiet listening conditions (Shannon et al., 1995). Periodicity cues include the representation of the fundamental frequency (f0) of the speaker's voice and carry the prosodic information.

The temporal fine-structure provides information about the spectrum and formant structure of speech sounds. Our understanding of how the human central auditory system processes temporal features in speech has been facilitated by neurophysiological studies that have examined central auditory coding of specific acoustic features present in speech sounds. This approach is based on the rationale that the speech signal is too complex to necessitate its decomposition into constituent components to describe neural mechanisms underlying each individual temporal feature (Abrams, 2017).

Given the complexity of the speech signal and well known non-linearity of the auditory system, the neural representation of isolated acoustic features present in speech cannot predict the representation of speech signal: the complex interaction of temporal features in the ongoing speech signal may result in different response characteristics than those predicted by isolated constituent parts.

Thus, one of the importance of using this evaluation is related to the hypothesis that attributes oral and/or written language disorders to a perceptual deficit, specifically to the auditory temporal processing.

The other importance is related to the relationship between cognitive function and auditory temporal abilities. There is no consensus, even today, about the actual effect of this aspect in auditory sensory tests. One study showed, for example, correlation between temporal tests and pitch discrimination thresholds with global intelligence measures, concluding that measures of temporal discrimination are related to specific aspects of intelligence.

Due to this lack of consensus, it is important to continue to study the auditory temporal processing related to speech and cognitive function, not to mention the importance of this ability regarding the hearing loss.

References

Abrams DA, Nicol T, White-Schwoch T, Zecker S, Kraus N (2017) Population responses in primary auditory cortex simultaneously represent the temporal envelope and periodicity features in natural speech. *Hearing Research*. 348:31-43

Filippini, R.; Schochat, E. (2014) A new paradigm for temporal masking assessment: pilot study. CoDAS. 26: 302-307.

Habib M. (2000) The neurological basis of developmental dyslexia. An overview and working hypothesis. Brain. 123 (12):2373–99

Moore BCJ. (2003) An introduction to the psychology of hearing. (eds.) USA: Academic Press.

Musiek FE. (1994) Frequency (pitch) and duration pattern tests. J Am Acad Audiol. 5(4):265- 8.

Musiek FE, Shinn JB, Jirsa R, Bamiou DE, Baran JA, Zaida E. (2005) GIN (Gaps-In-Noise) test performance in subjects with confirmed central auditory nervous system involvement. Ear Hear. 26(6):608-18

Rosen, S. (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. Philos. Trans. R. Soc. Lond B Biol. Sci. 336:367-373.

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

Shinn JB. 2003 Temporal processing: the basics. Hear J.56(7):52 2.

Tallal P, Piercy M. (1973) Developmental aphasia: impaired rate of non-verbal processing as a function of sensory modality. Neuropsychol. 11(4):389–98.

SYMPOSIUM:
LEVEL ANCHORED RATIO SCALING

Organised by Elisabet Borg.

SO WHAT'S THAT ON A SCALE FROM 1 TO 10?

Elisabet Borg
Department of Psychology, Stockholm University, SE-106 91 Stockholm, Sweden
eb@psychology.su.se

Abstract

The Borg CR Scales® are general psychophysical intensity scales intended for measurements of all kinds of perceptions and feelings. They are based on several important principles to give level-anchored ratio data with high interindividual validity. Some of these principles are: numerical coverage of the total subjective dynamic range; quantitative semantics for finding and choosing the best verbal (or pictorial) anchors to obtain high agreement concerning quantitative interpretation and preciseness for valid level-determinations; magnitude estimation for ratio data; congruence between numbers and anchors; Gunnar Borg's Range Model for interindividual comparisons; and a general point of reference or "fixed star" as a unit (for example a maximal perceived exertion). In this presentation, a short review is given of the rationale behind the Borg RPE scale® and Borg CR Scales®, and some examples of applications with the Borg centiMax Scale® (CR100) are given.

The oldest kinds of measurements are those of length and weight. When people began to trade their agricultural products on the market, it became important to be able to measure the weight of, e.g., a basket of grain and the silver that was used as payment. When they started to build houses, they needed to measure the length of a log of timber. In ancient Egypt and Mesopotamia standards of weight and length were created and have been dated as far back as 4-5000 years ago. For example, a standard length for the royal Egyptian cubit has been found from 2.500 BC. This was then divided into 7 hand widths and 28 finger widths and should represent the Pharaoh's forearm plus a hand width (NE, 2018). Like in this example, many scales and measures used "meaningful" units (forearm and hand width). Nowadays the standard units of the natural sciences have become much more sophisticated. Not many of us know the definition of the standard meter by heart ("the length of the path travelled by light in vacuum during a time interval of 1/299 792 458 of a second", BIPM, 1983).

Reliable and valid measurement is crucial for accurate conclusions about studied objects and phenomena. In psychology, what is being measured is, however, somewhat more complex than a basket of grain or a log of timber. People's behaviors, qualities, perceptions and emotions are uncertain, variable, and subjective. To develop scales so that obtained scores truly represent the individuals' characteristics is therefore a challenging task. The areas within psychology devoted to these questions are found within psychometrics and psychophysics. And while psychometrics primarily has focused on the development of items to study for example intelligence or personality, psychophysics has studied the properties of the numerical scales to study sensations and perceptions.

Stevens (1946) definition of measurement – "we may say that measurement, in the broadest sense, is defined as the assignment of numerals to objects or events according to rules" – broadened the possibility of scale construction within psychology. As long as there is a construct that can be operationally defined, what is required is a procedure that ensures that the assigned scores represent the studied phenomenon reliably and validly.

As a consequence, there are now hundreds, if not thousands, of psychological and health related tests and questionnaires, to measure everything from intelligence, pain, depression, etc., to satisfaction with the cashier at the pharmacy. And every such measuring instrument has a unique number of items combined with many kinds of (usually) numerical "rating scales" so that we end up with being surrounded by more or less meaningless numbers describing us and our achievements, characteristics, qualities and clinical symptoms... Jenny, for example, may have the following numbers attached to her: she is 174, 135, 17, 56, and 38. If I give you the additional information that these numbers give us information about her length, her intelligence, her depression, her stress level, and her hypochondria, you will probably guess correctly that she is 174 cm tall, and has an IQ of 135, and because you have preconceptions about both the scales and the variables you will make an interpretation and conclude that Jenny is rather tall and very smart. However, what do you understand from the other values? Probably not so much. And the reason is, of course, that the tests used for these kinds of variables all have a different number of questions and the rating-scales used for each question vary between the tests.

Ok, –so how clever is that on a scale from 1 to 10?

Another issue concerns the obtained data level. On scales like the one implied above from 1 to 10 (or often only 1 to 5) the data level can, at best, be assumed to be interval data. In many occasions, however, it would be fairer to say that only rank order, or ordinal data, is obtained. The information one can get from the data will only be "much" or "little", "more" or "less" of the variables measured (e.g., Svensson, 2000; G. Borg & E. Borg, 2001).

Psychophysical scaling and magnitude estimation

Stevens made direct ratio scaling, such as magnitude estimation, popular during the middle of the last century and the method was used to study a large number of sensory modalities (loudness, taste, brightness, heaviness, etc.). The method was usually carried out either with a standard stimulus that was assigned some convenient modulus (e.g. 10, or 100) or without a standard so that the observer "merely assigns whatever numbers seem appropriate to describe [for example] the loudness of a series of intensities presented in irregular order". One important idea with magnitude estimation was also to not impose any restrictions on the observer, who should be free to "select... quantifiers without constraint or prejudice" (Stevens, 1956). For most modalities, non-linear relationships were obtained that were described by power functions (e.g., Stevens, 1975). That magnitude estimation is measurement on a ratio level was supported by, e.g., Luce & Krumhansl (1988), Luce (1990), and Naarens (1996).

Because of the success with ratio scaling, Stevens was hoping that traditional categorical rating-scales would gradually be replaced by magnitude estimation, but in 1971 he wrote that "in almost two decades of practice, magnitude estimation has not displaced category estimation — nor does it seem likely to do so any time soon." One reason is, as my father Gunnar so often has pointed out, "the fact that the ratio methods only give relative intensities and no subjective "levels" for immediate interindividual or intermodal comparison." On the other hand, this is obtained with category scales that use well defined verbal labels (or pictures) to describe the intensity levels, such as "weak" and "strong". An interesting question was therefore how to combine the advantages of both scale types (G. Borg, 1982, 1998).

Gunnar Borgs' range model for interindividual and inter-process comparisons

To be able to make comparisons between individuals (or groups of individuals) an interindividually valid framework for comparisons, is needed. Biological differences in sensory functioning or capacity, as well as differences in experiences and cognitive functioning, will, for example, always exist. However, because all biological systems have limited capacity, from the threshold to an upper limit, Gunnar Borg argued that the perceptual range from min to max could be set equal for all subjects. By additionally using a reference level with a high degree of interindividual agreement as a kind of "fixed star", interindividual and inter-process comparisons are possible. A previously experienced maximal exertion, of for example lifting something that is close to maximally heavy, has been found to work well for this purpose (G. Borg, 1962, 1990, 1992; Sagal & Borg, 1993).

Principles for scale construction

A good measuring instrument for quantitative psychological variables should be valid, and for example, have high content and construct validity. But, it is also important that 1) the numerical range covers the total possible perceptual range for the studied construct, 2) data can be treated as ratio data, similar to what is obtained with magnitude estimation (or at least interval data), 3) anchors, weather verbal or pictorial, are carefully chosen and placed on the scale so that congruence is obtained between anchors and numbers, 4) two-way communication is facilitated, 5) end-effects are avoided, 6) interindividual and inter-process comparisons are made possible, and that 7) if available, good use is made of physiological, or other, correlates (G. Borg & E. Borg, 2001).

The Borg RPE scale®

The 6-20 scale for ratings of perceived exertion (RPE) was constructed as an interval scale. The number range was chosen to cover the natural range of a heart rate variation from 60 - 200 bpm for work on a bicycle ergometer with a stepwise increase of the workload every 2-3 minutes. Interval data was obtained by choice of the anchors and how they were placed on the numerical scale (G. Borg, 1962, 1970, 1998). The scale has been much appreciated for its simplicity and usefulness in a wide range of situations, for exercise testing, physical training and rehabilitation.

The Borg CR scales®

The idea with the level anchored ratio scales was to place verbal anchors with high interpersonal correspondence in quantitative meaning ("interpretation") and low standard deviation ("preciseness") on a numerical scale so that ratio data is obtained. In order to do so the relationship found between a category scale and a ratio scale was used to find the best numerical position of the anchors on the ratio scale. (Stevens & Galanter, 1957, G. Borg, 1982, 1998). Thus, the scale becomes a Category-Ratio (CR) scale (see also, E. Borg, 2011).

The most well-known CR scale, is the Borg CR10 Scale®, covering an intensity variation from "Nothing at all" (=0) to "Extremely strong, Maximal" (=10), and with a possibility to use values somewhat above 10 ("Absolute maximum", = •) in extreme situations. The scale is continuous and decimals should be allowed. The reliability and validity has been shown in many applications and for a manifold of different sensory modalities, but it's primary usage has been for perceived exertion, dyspnea and pain perception. (For some examples, see e.g., G., Borg, 1998, E. Borg, 2007).

Because the CR10 scale was somewhat rough, a more finely graded CR scale has also been developed, the Borg centiMax Scale® (CR100). Like the CR10 scale, the verbal anchors are chosen so that the scale can function as a general intensity scale, and it is constructed to give ratio data that grow linearly with magnitude estimation (Fig. 1). Because the scale values are given as centigrade with "Maximal" (= 100) defined as a previously experienced maximal exertion, the scale is called a centiMax scale. This is therefore also the scale unit and can be written "cMax" or "cM" for short. (G. Borg & E. Borg, 2001; E. Borg, 2007).

Some applications with the Borg centiMax Scale® (CR100)

The main areas of application for the CR scales are still to be found for perceived exertion, fatigue, breathlessness, pain, etc., but especially because the centiMax scale is more finely graded this scale has also been used in several new areas the last few years. Some more recent examples are: ratings of perceived exertion in sports (Fanchini, Ferraresi, Modena, Schena, Coutts, & Impellizzeri, 2016; Fanchini, et al., 2017; Ferrari, Chirico, Rasá, 2016; Weston, Siegler, Bahnert, McBrien, & Lovell, 2015); putting force in golf (Molander, Boraxbekk, Stenling, & Borg, 2015); perceived exertion and breathlessness in medicine and rehabilitation (E. Borg, G. Borg, Larsson, Letzter, & Sundblad, 2010; Keser, Suyani, Yosmaoglu, Aki & Turkoz Sucak, 2014; Lazarinis et al., 2014); for endurance performance (Fabre, Mourot, Zerbini, Pellegrini, Bortolan, & Schena, 2013); but also for perceived intensity of facial expressions (Gerhardsson, Högman, & Fischer, 2015); in continuous scaling of aircraft noise (Dickson & Bolin, 2014); in olfactory research (Olofsson, Niedenthal, Ehrndal, Zakrzewska, Wartel, & Larsson, 2017); for depression (E. Borg & Sundell, 2017; Magalhães, 2017); and for hypernasality ratings (Yamashita, E. Borg, Granqvist, & Lohmander, 2018). A couple of examples will be presented in the following parts in this symposium.

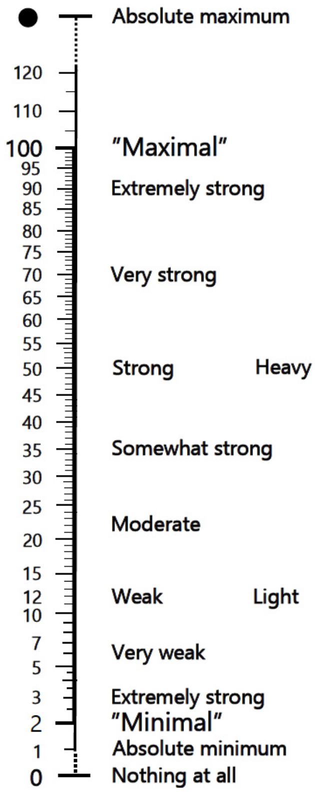


Fig. 1. The Borg centiMax Scale® (CR100). © G. Borg & E. Borg, 2001, E. Borg, 2007.

References

Borg, E. (2007). *On Perceived Exertion and its Measurement (Doctoral Dissertation)*. Stockholm, Sweden: Department of Psychology, Stockholm university. Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-6862

Borg, E. (2011). Placing verbal descriptors on a ratio scale. In D. Algom, D. Zakay, E. Chajut, S. Shaki, Y. Mama, & V. Shakuf (Eds.), *Proceedings of the 27th Annual Meeting of the International Society for Psychophysics* (pp. 119–124). Raanana, Israel: The International Society for Psychophysics. Retrieved from http://www.ispsychophysics.org/fd/index.php/proceedings/article/view/401

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), 644–50. http://doi.org/10.1111/j.1600-0838.2009.00985.x

Borg, E., & Sundell, J. (2017). Scaling depression with psychophysical scaling: A comparison between the Borg CR Scale® (CR100, centiMax®) and PHQ-9 on a non-clinical sample. In K. Ueda, H. Ito, G. B. Remijn, & Y. Nakajima (Eds.), *Proceedings of the 33rd Annual Meeting of the International Society for Psychophyiscs* (pp. 101–107). Fukuoka, Japan: International Society for Psychophysics.

Borg, G. (1962). *Physical performance and perceived exertion.* Studia Psychologica et Paedagogica, Series altera, Investigationes XI. Gleerup.

Borg, G. (1970). Perceived exertion as an indicator of somatic stress. *Scandinavian Journal of Rehabilitation Medicine*, *2*(2), 92–98. http://doi.org/S/N

Borg, G. (1982). A category scale with ratio properties for intermodal and interindividual comparisons. In H.-G. Geissler & P. Petzold (Eds.), *Psychophysical Judgment 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. Hezewijk, & S. Terwee (Eds.), *Recent Trends in Theoretical Psychology* (Vol. 2, pp. 439-444). New York: Springer-Verlag.

Borg, G. (1992). A "fixed star" for interprocess comparisons. In *Fechner Day 92. Proceedings of the Eigth Annual Meeting of the International Society for Psychophysics.*

Borg, G. (1998). *Borg's perceived exertion and pain scales*. Champaign, IL: Human Kinetics.

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

CGPM, (1983), *Comptes Rendus de la 17^e^ CGPM* (1983), 1984, p.97 https://www.bipm.org/en/CGPM/db/17/1/ (retrieved 2018-05-24)

Dickson, C., & Bolin, K. (2014). Continuous judgment by category-ratio scaling of aircraft noise. *Applied Acoustics*, *84*, 3–8. http://doi.org/10.1016/j.apacoust.2014.02.013

Fabre, N., Mourot, L., Zerbini, L., Pellegrini, B., Bortolan, L., & Schena, F. (2013). A novel approach for lactate threshold assessment based on rating of perceived exertion. *International Journal of Sports Physiology and Performance*. http://doi.org/10.1123/ijspp.8.3.263

Fanchini, M., Ferraresi, I., Modena, R., Schena, F., Coutts, A. J., & Impellizzeri, F. M. (2016). Use of CR100 scale for session-RPE in soccer and interchangeability with CR10. *International Journal of Sports Physiology and Performance*. http://doi.org/10.1123/ijspp.2015-0273

Fanchini, M., Ferraresi, I., Petruolo, A., Azzalin, A., Schena, F., Impellizzeri, F. M., Schena, F. (2017). Is a retrospective RPE appropriate in soccer? Response shift and recall bias. *Science and Medicine in Football*. http://doi.org/10.1080/02640414.2016.1231411

Ferrari, G., Chirico, F., & Rasá, G. (2016). Examining physical training versus physical and mental training programmes in Swimrun semi-professional athletes: A randomised, controlled, trial. *Journal of Health and Social Sciences*, *13*, 199–210. http://doi.org/10.19204/2016/gndr22

Gerhardsson, A., Högman, L., & Fischer, H. (2015). Viewing distance matter to perceived intensity of facial expressions. *Frontiers in Psychology*. http://doi.org/10.3389/fpsyg.2015.00944

Keser, I., Suyani, E., Yosmaoglu, H. B., Aki, S. Z., & Turkoz Sucak, A. G. (2014). Acute physiological responses to physiotherapy applications pre and post autologous stem cell transplantation: an experimental study. *Hematology*, *19*(3), 136–140. http://doi.org/10.1179/1607845413Y.0000000106

Lazarinis, N., Jørgensen, L., Ekström, T., Bjermer, L., Dahlén, B., Pullerits, T., … Larsson, K. (2014). Combination of budesonide/formoterol on demand improves asthma control by reducing exercise-induced bronchoconstriction. *Thorax*, *69*(2), 130–136. http://doi.org/10.1136/thoraxjnl-2013-203557

Luce, R. D. (1990). "On the Possible Psychophysical Laws" Revisited: Remarks on Cross-Modal Matching. *Psychological Review*, *97*(1), 66–77. http://doi.org/10.1037/0033-295X.97.1.66

Luce, D. R., & Krumhansl, C. L. (1988). Measurement, scaling and psychophysics. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Stevens' handbook of experimental psychology: Perception and motivation; Learning and cognition* (pp. 3–74). Oxford, England: Wiley & Sons. http://doi.org/10.1016/0001-6918(90)90085-T

Magalhães, A. (2017). *Três estudos usando a Escala centiMax de Borg® (Escala CR de Borg®, CR 100, cM) para o escalonamento de sintomas depressivos.* (PhD Thesis), Univesity of São Paulo, São Paulo.

Molander, B. O., Boraxbekk, C., Stenling, A., & Borg, E. (2015). Judging and Regulating Force in Sports: The Psychophysical Scale as a Tool for Putting Shots a Tool for Putting Shots Judging and Regulating Force in Sports. *Journal of Sport Psychology in Action*, *6*(3), 154–169. http://doi.org/10.1080/21520704.2015.1092484

Narens, L. (1996). A theory of ratio magnitude estimation. *Journal of Mathematical Psychology*, *40*(2), 109–129. http://doi.org/10.1006/jmps.1996.0011

Nationalencyklopedin, mätning. http://www.ne.se.ezp.sub.su.se/uppslagsverk/encyklopedi/lång/mätning (retrieved 2018-05-24)

Olofsson, J. K., Niedenthal, S., Ehrndal, M., Zakrzewska, M., Wartel, A., & Larsson, M. (2017). Beyond Smell-O-Vision: Possibilities for Smell-Based Digital Media. *Simulation and Gaming*, *48*(4), 455–479. http://doi.org/10.1177/1046878117702184

Sagal, P., & Borg, G. (1993). The range principle and the problem of other minds. *British Journal for the Philosophy of Science*, *44*(3), 477–491. Retrieved from http://www.jstor.org/stable/688017

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*, 677-680.

Stevens, S. S. (1965). The Direct Estimation of Sensory Magnitudes: Loudness. *The American Journal of Psychology*, *69*(1), 1–25.

Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. *New York Wiley*. http://doi.org/10.2307/1421904

Stevens, S. S., & Galanter, E. H. (1957). Ratio Scales and Category Scales for a dozen perceptual continua. *Journal of Experimental Psychology*, *54*(6).

Svensson, E. (2000). Concordance between ratings using different scales for the same variable. *Statistical Medicine*, 19, 3483–3496.

Weston, M., Siegler, J., Bahnert, A., McBrien, J., & Lovell, R. (2015). The application of differential ratings of perceived exertion to Australian Football League matches. *Journal of Science and Medicine in Sport*. http://doi.org/10.1016/j.jsams.2014.09.001

Yamashita, R. P., Borg, E., Granqvist, S., & Lohmander, A. (2018). Reliability of Hypernasality Rating: Comparison of 3 Different Methods for Perceptual Assessment. *The Cleft Palate-Craniofacial Journal*, (online april 2018). http://doi.org/10.1177/1055665618767116

ASSESSING TRAITS IN A PSYCHOPHYSICAL WAY: REASSESSING NEED FOR COGNITION AND BEHAVIORAL INHIBITION/APPROACH

Sebastian Cancino-Montecinos and Elisabet Borg
Department of Psychology, Stockholm University, SE-106 91 Stockholm, Sweden
sebastian.cancino@psychology.su.se, eb@psychology.su.se

Abstract

The aim of this study was to investigate if different scale formats affect what conclusion one can draw about the prevalence of a specific trait in a sample. More specifically, we compared the original scale format of Need for cognition (1-5) and Behavioral Inhibition/Approach (1-4) with an 11-point scale (0-10), and a psychophysical scale originally developed to measure physical exertion, Borg centiMax Scale®. Forty-eight psychology undergraduate students participated in return for course credit. In a within-subjects design, all participants completed both questionnaires in all three versions. Results revealed that the mean was consistently reaching ceiling effects when using the original scale formats, and the variation was relatively low compared to the other scales. In sum, the results revealed that the scale format plays a significant role in how prevalent a specific trait becomes in a sample.

The main focus in psychophysics is to investigate the relationship between physical stimuli and people's perception of those *stimuli (*Gescheider, 1997*). In* psychometrics, the main focus is to investigate individual differences in traits (e.g., personality, intelligence etc.) among people (Kaplan & Saccuzzo, 2010). These different themes of psychological inquiry are seemingly quite distant from each other. However, they do share a common goal. In both psychophysics and psychometrics, the issue of measurement is at the center point. Note however that the issue of measurement is twofold. First, the researcher must develop adequate equipment in order to capture the psychological, or physiological, phenomena correctly. In psychometrics, the adequate equipment is a questionnaire containing a very specific set of items. Second, the researcher must also develop a response format that reliably measures the phenomena at hand. In a questionnaire, the ratings are usually done on a Likert scale ranging from 1-5 or 1-7, were all increments are treated as equally large (Furr, 2011). These numerical units are often coupled with sematic labels, for instance: 1 = Strongly disagree; 2 = Disagree; 3 = Neither agree nor disagree; 4 = Agree; 5 = Strongly agree. In our view, here lies the problem. Since, psychometricians assume equal distance between numbers, the semantic labels must correspond with this assumption. That is, there must be equal distance between the sematic meaning of, for instance, strongly disagree, disagree, and neither agree nor disagree, etc. Usually, the correspondence between labels and numbers is never tested. Another problematic issue is that if the scale is too narrow (i.e., too few rating points), it might not do a good job at differentiating between individuals – which is the main purpose in psychometrics. In the case above, there might exist several semantic labels between strongly disagree and disagree, but also between disagree and neither agree nor disagree. Thus, a researcher might miss out on making a more fine-grained differentiation between individuals if the response format of the scale is not properly developed. In this study, we will compare how different response formats affect what conclusions one might draw about the prevalence of a particular trait in a sample (see also Dawes, 2008). Also, how these formats might affect the shape of the distributions. Note that we are *not* criticizing the psychological constructs. In fact, we do assume that constructs studied in this paper are real. The constructs we will investigate are *Need for cognition (NFC;* Cacioppo & Petty, 1982), and *Behavioral Inhibition/Approach* (*BIS/BAS;* Carver & White, 1994).

Scale formats

The NFC questionnaire is used to assess people's inclination for thinking and deliberating for long hours and finding enjoyment in doing so. The NFC uses a five-point scale with the following labels: 1 = Extremely uncharacteristic; 2 = Somewhat uncharacteristic; 3 = Uncertain; 4 = Somewhat characteristic; 5 = Extremely characteristic. The most obvious problem here is that value 3 is paired with the label *Uncertain*. This label is problematic because "uncertain" does not mean that the individual experiences that item to any specific degree. The word uncertain means rather that the individual is not sure about his/her opinion regarding the item. Clearly, this does not correspond with the numerical meaning; having degree of 3 on the five-point scale. A clear way of illustrating this problem is to imagine an individual scoring on average 3.0 (with little variance) on the NFC. If that individual follows the instructions of questionnaire, what does a score of 3.0 mean? Does it mean that the individual is uncertain about most of the items? Or does it mean that the individual has a need for cognition of 3 (out of 5)? That is, neither characteristic nor uncharacteristic. The second problem is the leap from 1-2 and 4-5. That is, the leap from *extremely* to *somewhat*, and *somewhat* to *extremely*, is rather large. We can think of at least two additional steps between *somewhat* and *extremely* – for instance, simply *Characteristic*, *and Very*. These two related problems could result in a scale that gives researchers ordinal data with two categories (Characteristic and Uncharacteristic) with two levels (Extremely and Somewhat), and a midpoint with a semantic label that poorly represents the numerical counterpart.

The BIS/BAS questionnaire was developed to assess individual differences in people's tendencies to inhibit and approach social and personal situations. One part of the questionnaire measures inhibition (*BIS*), while three other parts measure approach (*BAS-Drive*; *BAS-Fun Seeking*; *BAS-Reward Responsiveness*). The BIS/BAS has similar problems to the NFC, but also some unique issues. The scale of the questionnaire has the following steps (1) Very true for me; (2) Somewhat true for me; (3) Somewhat false for me; (4) Very false for me. Again, there is an overwhelming risk of receiving ordinal data. That is, two categories (true/false) with two levels (very/somewhat). When reading the original study of White and Carver (1994), there is no rationale for the format of the scale. Also, the words true and false are rather problematic in this context. These words are absolute, meaning that something cannot be "somewhat" true or false. Something is either true *or* false. Finally, a critical problem for both the NFC and the BIS/BAS, is that there is no natural congruence between the numerical values and the semantic labels. Some individuals might follow what the labels actually mean (e.g., "Uncertain"), while others might make an interpretation of what they believe the researcher means. In worst case, the scale has an in-built error that might systematically collect erroneous data.

The *Borg centiMax scale*® is a category scale with ratio scale properties and was originally developed to measure people's subjective experience of exertion (see G. Borg & E. Borg, 1987; 1994; 1998; 2001; 2002). The uniqueness in this scale lies in that the labels are carefully chosen for different levels of exertion. Thus, unlike traditional scaling methods, with the centiMax the researcher obtains information about what a particular numerical value means (psychologically) to the individual. Furthermore, unlike scales in personality and social psychology, the labels in the centiMax have been derived from testing, and these labels are placed were they, perceptually, belong on a ratio scale. Although the centiMax has not been used to measure personality or behavioral tendencies, we argue that this scale format (which is more precise and less confusing) might do a better at assessing the prevalence of individual differences.

The present study

In the present study, we will compare the original scales of the NFC (1-5) and BIS/BAS (1-4) with an 11-point scale (0-10) as a reference, and the centiMax (0-130). We predict that the data received from the 11-point scale format, and the centiMax will resemble interval data to a higher degree compared to the 4-point (BIS/BAS) and 5-point scale format (NFC). Also, because of the relatively few rating points, the original scales will consistently have higher mean values and less variation compared to the 11-point scales and the centiMax (see e.g., Dawes, 2008).

Method

Participants

Forty-eight students, mean age was 24.2 (*SD* = 5.7), and 41 were women, were recruited for this study, and they received extra course credit after completing the study.

Procedure and materials

There were three different scale formats: (1) the original format; (2) the 11-point format; and (3) the centiMax format. To avoid order-effects, the questionnaires were randomly assigned with a three-day interval between each session. In all sessions, participants were first asked to complete both questionnaires (NFC; 18 items, and BIS/BAS; 21 items), and were later asked to assess how difficult it was to respond to each item. All sessions were administrated online. Note that the scale format of the BIS/BAS was reversed when performing the analyses. That is, the more "true" each item was to the individual, the higher the numerical value. This makes it easier to compare the scale formats.

Results

Distributions

Results show that skewness does not vary considerably depending on scale format (see Figure 1). The biggest difference was found in the BAS-Drive, were the centiMax scale was positively skewed. The same is true for kurtosis, across the different scale formats none of them differ noticeably from each other. However, the figure shows that the distributions with the original scale formats are consistently closer to ceiling effects compared to other scale formats. Finally, the internal consistency was higher (or equally high) with the centiMax scale format compared to both the original and 11-point format (all α's > .75).

Further comparisons between scale formats

The second set analysis aimed at further investigating the effects of scale format. First, we compared the original scale with the centiMax by using the 11-point scale as a reference point. This was done the up-scaling the original scale and down-scaling the centiMax so they corresponded with the 11-point scale. In the NFC, the 11-point scale is 2.2 times larger than the original, and 11.82 times smaller than the centiMax. Thus, new variables were created by multiplying the means obtained with the original scale by 2.2, and by multiplying the means of the centiMax by .08461538 – which is the ratio of the 11-point scale and the centiMax. The same procedure was performed for the BIS/BAS. However, in this case, the means of the original scale were multiplied by 2.75, and again by .08461538 for the means of the centiMax. Results reveal that all means (relative the means obtained by the 11-point scale) increased after up-scaling the original scale to the 11-point scale, and all means decreased after down-scaling the centiMax to the 11-point scale. Also, the variation, across all constructs, was consistently larger with the centiMax compared to the other scale formats (see Table 1, 2, 3).

Discussion and conclusions

It is important to use a scale format that gathers data in a correct and comprehensive way. By this, we mean that a scale format must be constructed in such a way that it is able to measure the natural occurrence of the trait as closely as possible. A scale format that is too restricted and coupled with confusing semantic labels (e.g., 1 = very true for me; 4 = very false for me) might not capture the full range (and variation) of the trait. Or even worse, it might give researchers erroneous data. This study shows that the scale format will have an effect on the mean, standard deviation and frequency distribution. In short, the original scale formats of the NFC and BIS/BAS consistently (a) reached ceiling effects, and (b) had less variation compared to the centiMax scale. The ceiling effects problem with the original scales does speak for the use of a centiMax scale format, since it might not be reasonable to expect these effects given that the traits investigated in this study are thought to be normally distributed. In sum, having verbal expressions that naturally correspond with numerical values truly do affect what scale value participants receive (Borg, G.& Borg, E, 2001), and leaves possibly less room for measurement error.

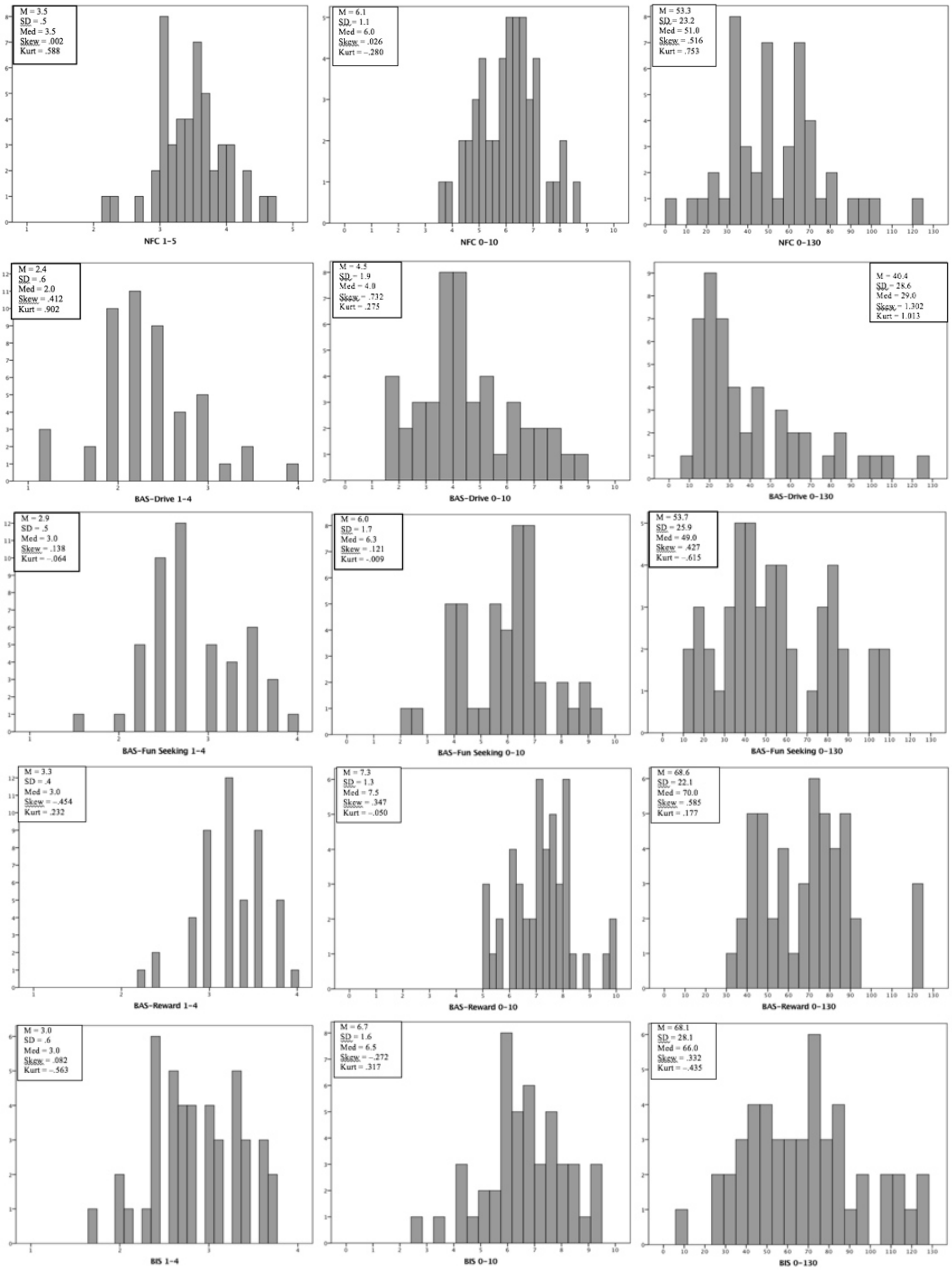


Figure 1. Distribution, mean, standard deviation, median, skewness and kurtosis for NFC and BIS/BAS across scale format.

Table 1. Mean, standard deviation, mean difference and Cohen's d for the 11- point scale compared to the up-scaled original scale.

	M (SD)- 0-10	M (SD)-Original up-scaled to 0-10	Mean difference	Cohen's d
NFC	6.1 (1.1)	7.6 (1.1)	1.5	1.4
BAS-Drive	4.5 (1.9)	6.6 (1.5)	2.1	1.2
BAS-Fun Seeking	6.0 (1.7)	7.9 (1.4)	1.9	1.2
BAS-Reward	7.3 (1.3)	8.9 (1.1)	1.6	1.4
BIS	6.7 (1.6)	8.1 (1.6)	1.4	.9

Table 2. Mean, standard deviation, mean difference and Cohen's d for the 11- point scale compared to the down-scaled centiMax.

	M (SD)- 0-10	M (SD)-centiMax down-scaled to 0-10	Mean difference	Cohen's d
NFC	6.1 (1.1)	4.5 (1.9)	1.6	1.0
BAS-Drive	4.5 (1.9)	3.4 (2.4)	1.1	.5
BAS-Fun Seeking	6.0 (1.7)	4.5 (2.2)	1.5	.8
BAS-Reward	7.3 (1.3)	5.8 (1.9)	1.5	1.0
BIS	6.7 (1.6)	5.8 (2.4)	.9	.5

Table 3. Mean, standard deviation, mean difference and Cohen's d for the up-scaled original scale compared to the down-scaled centiMax.

	M (SD)-Original up-scaled to 0-10	M (SD)-centiMax down-scaled to 0-10	Mean difference	Cohen's d
NFC	7.6 (1.1)	4.5 (1.9)	3.1	2.0
BAS-Drive	6.6 (1.5)	3.4 (2.4)	3.2	1.6
BAS-Fun Seeking	7.9 (1.4)	4.5 (2.2)	3.4	1.8
BAS-Reward	8.9 (1.1)	5.8 (1.9)	3.1	2.0
BIS	8.1 (1.6)	5.8 (2.4)	2.3	1.2

In a larger perspective, this might affect what conclusion one draws about the prevalence of a particular trait in a population, and also how this trait is distributed across the population (standard deviation, median, mode etc.). These conclusions might further influence decisions made by policy makers or employers. For instance, if an employer is selecting workers based on their natural tendencies to socialize and converse with people, he/she might think that someone with an average of 2.3 does not differ so much from someone with average of 2.7. However, if these results were obtained using a scale format resembling the one used in the BIS/BAS questionnaire, one individual's average is approximately "somewhat false", and another individual's average is approximately "somewhat true". The obvious problem here is that the words true and false have no natural position on a quantitative continuum. Thus, the employer might end up with an individual reluctant to socialize and converse with people – which is the opposite of what he/she was looking for. This example clearly shows the problem when sematic labels do not correspond with the numerical value. We conclude that the centiMax format does a better job at assessing the prevalence of traits, even though this scale was developed to measure physical exertion. Future research should try to develop scale formats, using the centiMax rationale, that makes sense for the particular psychological construct one is trying to measure.

References

Borg, E., and Borg, G. (2002) A comparison of AME and CR100 for scaling perceived exertion. *Acta Psychologica*, *109*, 157–175.

Borg, G. (1998). *Borg's perceived exertion and pain scales*. Champaign, IL: Human Kinetics

Borg, G., and Borg, E. (1994). Principles and experiments in Category-Ratio scaling. *Reports from the Department of Psychology*, Stockholm University No. 789.

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

Borg, G., and Borg, E. (1987). Need for an additional constant in the psychophysical function of taste perception. *Reports from the Department of Psychology*, Stockholm University No. 668.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42,* 116–131.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology*, 67, 319–333.

Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. International Journal of Market Research, 50, 61–77.

Furr, R. M. (2011). *Scale construction and psychometrics for social and personality psychology*. Thousand Oaks, CA: Sage.

Gescheider, G. (1997). *Psychophysics: the fundamentals (3rd ed.). Lawrence Erlbaum Associates.*

Kaplan, R.M., & Saccuzzo, D.P. (2010). *Psychological Testing: Principles, Applications, and Issues.* (8th ed.). Belmont, CA: Wadsworth, Cengage Learning.

Likert, R. (1932). *A Technique for the Measurement of Attitudes. Archives of Psychology, 140: 1–55.*

RELIABILITY OF AUDITORY PERCEPTUAL ASSESSMENT OF HYPERNASALITY IN SPEECH USING DIFFERENT SCALES

Anette Lohmander[1], Renata Yamashita[2], Svante Granqvist[3], and Elisabet Borg[4]
[1]*Division of Speech and Language Pathology, Department of Clinical Science, Intervention and Technology, Karolinska Institutet, SE-14186, Stockholm Sweden.*
[2]*Laboratory of Physiology, Hospital for Rehabilitation of Craniofacial Anomalies, University of São Paulo, Bauru, Brazil.*
[3]*Royal Institute of Technology, Stockholm, Sweden*
[4]*Department of Psychology, Stockholm University, SE-106 91 Stockholm, Sweden*

Auditory-perceptual assessment has been criticized because of its inherent subjectivity. However, perceptual judgments are the primary tool in the clinical assessment of voice and resonance parameters and a key task for the speech–language pathologist/therapist (SLP/T). Decision on treatment or further examination is based on the perceptual speech assessment which need to be reliable. Of all perceptual dimensions used to distinguish normal from abnormal speech the most difficult to judge reliably is hypernasality (Watterson et al., 2017). Nevertheless, the final decision regarding whether an individual has nasality or other speech problems is based on the listener's subjective measurement (Moll, 1964). Nasality is present in normal voice production and refers to perceived nasal sounds arising from the coupling of the oral and nasal resonating cavities. Nasal consonants are common in the languages. Normal nasal resonance has a range of acceptability and is perceived along a continuum, while nasal resonance disorders are associated with defects or dysfunction in the palate. In the current study the reliability of perceptual assessment of hypernasality with three different methods were compared. Standardised audio recordings of 5-year-old Swedish-speaking children with repaired cleft palate consisting of 73 stimuli in three different randomised orders were perceptually assessed by four experienced speech-language pathologists using three different methods: a sort and rate procedure (VISOR) allowing comparison between and ordering stimuli along a visual analogue scale; a 2-step method beginning with determination whether the speech resonance is within normal range or not. If not, the stimulus is rated along a 3 point ordinal category scale; a combined category-ratio scale (Borg centiMax®(cM)), where verbally well-defined categories are level-anchored to a ratio scale. Each listener completed a total of 657 hypernasality ratings (73 stimuli in 3 orders by 3 methods) for the rating task. Good to excellent intra-rater reliability was found within each listener for all methods. The highest inter-rater reliability was demonstrated for VISOR and the Borg cM. High consistency within each method was found with the highest for the Borg cM. In conclusion, both the VISOR and the Borg centiMax® seem appropriate for auditory-perceptual assessment of hypernasality with similar and high reliability. However, the Borg cM showed slightly better consistency and seems easiest to use in a clinical setting. Further research should aim for valid definitions of how hypernasality corresponds to the categories along the scale.

References

Moll, K. L. (1964). 'Objective' measures of nasality. *Cleft Palate Journal*, 1, 371–374.

Watterson, T., Lewis, K., Allord, M., Sulprizio, S., & O'Neill, P. (2007). Effect of vowel type on reliability of nasality ratings. *Journal of Communication Disorders*, 40, 503–512.

Yamashita, P. R., Granqvist, G., Borg, E., & Lohmander, A. Reliability of hypernasality rating: comparison of three different methods for perceptual assessment. *Cleft Palate-Craniofacial Journal*. Ahead of Print 2018.

IMPROVING THE PSYCHOLOGICAL ASSESSMENT OF PATIENTS BY USING PSYCHOPHYSICAL METHODS

Adsson Magalhães[1], Elisabet Borg[2], Ewa Mörtberg[2], and Marcelo Costa[1]
[1] *Department of Experimental Psychology, Institute of Psychology, University of São Paulo, São Paulo, Brazil*
[2] *Department of Psychology, Stockholm University, Stockholm, Sweden*
adsson.magalhaes@usp.br

Abstract

For many years, psychologists have developed instruments for assessing disorders based on ordinal scales. These scales, despite being widely used, easy to apply, and producing good results, have some imperfections. By bringing the knowledge acquired by the psychophysics field during the last century, it is possible to improve psychological assessment creating new instruments based in psychophysical methods of measurement like ratio scaling.

Measurement is a term that brings forward questions regarding reliability, precision, technique, and bias (Luce & Krumhansl, 1988). According to Ferris (2004, p. 107), "measurement is an empirical process, using an instrument, effecting a rigorous and objective mapping of an observable into a category in a model of the observable that meaningfully distinguishes the manifestation from other possible and distinguishable manifestations". For both, psychological and physical attributes, our sensations suggest a measurement system that codes the degree or amount of the attribute or property in question. In Psychology, we are concerned with the sensations themselves, and this difference poses a challenge to develop measurements that are appropriate for our special purposes (Luce & Krumhansl, 1988).

The fact that sensations can't be separated into subparts, or placed side-by-side on a ruler, was one of the arguments used against them being measured. But that was before there was a general recognition that measurement is not limited to counting (Stevens, 1975). Stevens (1959) described that nominal, ordinal, interval, and ratio are the most important empirical properties of the world for which numbers can serve as models. The question of how we can transform numbers without losing information from the results of empirical experiments, can be answered by using ratio scales as they rely on the conception of invariance.

Assessment in Psychology

Because there are different techniques to assess psychological attributes, this is one of the domains of practical Psychology that has brought discussion since the beginning. Used as parallel resources to DSM (APA, 2013) and ICD (WHO, 1992), psychological instruments aim to measure the intensity of symptoms of disorders like anxiety and depression. They are able to assess someone's self-perception and are valuable help for health professionals to correctly diagnose and analyze characteristics that are hard to observe clinically or by interview.

Psychologists are interested in a wide variety of behaviors in humans and animals, and the interests of the investigator determine, to a large extent, which of the observable behaviors receive attention. Once this choice has been made, the investigator faces the problem of finding the best way to measure these behaviors, perceptions, and attributes (Luce & Krumhansl, 1988).

The application of psychophysical assessments in clinical settings may also provide empirical elements for the study of epistemological issues of perception, the investigation of how the physical world is perceived by patients with different diseases such as cerebral palsy, schizophrenia, autism, depression, among others (Costa, 2011). However, some problems appear regarding the precision and reliability of the instruments.

Problems with Measurement in Psychology

Taking the example of measures for Depression, the assessment forms exiting usually consist of questions to measure the symptoms and severity of depression with a simple categorical scale (Likert type), like Beck Depression Inventory – BDI (Beck *et al.,* 1961), Hamilton Rating Scale for Depression – HAMD (Hamilton, 1960) and Montgomery–Åsberg Depression Rating Scale - MADRS (Montgomery & Åsberg, 1979), etc.

There is a weakness present in these instruments. The sentences (stimuli) are ordinal based and it is not always true that distances between the stimuli that are measured can be ordered using this kind of scale (Stevens, 1958). This reflects directly in restricting the possibilities of mathematical and statistical analysis.

An example found in studies is that simply the presence or absence of a symptom may not be an enough reliable indicator of depressive disorders (Schwenk, Coyne & Fechner-Bates, 1996). There are situations in which relative distances are wrongly ordered. In many instruments of psychological assessment, for one item the difference between two steps on the scale (e.g., between 0 and 1) is different from the difference between the same two steps (0 and 1) for another item. An illustration of this is found from comparing two BDI questions: the “zero” for Past Failure (item 3 – I do not feel like a failure) with “zero” in Guilt (item 5 – I don’t feel particularly guilty). Obviously the first statement corresponds closely to “Nothing at all” when it comes to the intensity of the feeling of past failure, whereas the second statement – Not feeling “particularly” guilty – means a week feeling of guilt (somewhat guilty, but not so strong?). These two “zeroes” are thus not equivalent in their perceptual “meaning”.

In a psychophysical context, the weaknesses are rather related to statistical aspects. According to Hadžibajramović, Svensson and Ahlborg Jr. (2013), some variables are measured by multi-item questionnaires, in which each item usually has several numerically coded response categories. These values are rank ordered, meaning that each category has more of the attribute being measured than the previous category, but the differences between the categories are unknown and thus these values don’t have the mathematical properties needed for arithmetic calculations. The definition of meaningful rules for the measurement of constructs like depressive symptoms, depends on the application field, the paradigm, and the measurement theory, as well as statistical knowledge, which may be better achieved by psychophysical scaling.

Scaling methods that assess the perception of time in depressive subjects are already used in studies of depression (Kornbrot, Msetfi & Grimwood, 2013), but as described above, there is a need for better instruments for screening and evaluation of Major Depressive Disorder. Only the ratio scale maps numbers to sensations, so no mathematical transformation (such as multiplication by a constant) causes loss of information and permits specification of when a feeling is X times more intense than another (Stevens, 1955, 1958a, 1958b, 1959).

With a new perspective brought from psychophysical scaling, a scrutiny of the symptoms could be made in a better way. With ratio data symptom profiles are possible and may, for example, convey interesting information about the relative importance of specific symptoms for separate individuals (G. Borg, 1998, E. Borg, 2007).

The Borg centiMax Scale® (CR100)

In order to develop a new scale, Prof. Dr. Gunnar Borg (e.g., 1962, 1998) started a process of combining labeled category scaling with ratio scaling to obtain a level anchored ratio scale, that provides basic empirical operations such as determinations of equality of ratio and mathematical transformations. The latest version of the Borg CR Scales is The Borg centiMax Scale® (CR100) (G. Borg & E. Borg, 2016). Ranging from 0 to 100, it has verbal anchors placed in accordance with their perceptual intensity along the numerical scale, generating ratio data. In general, the verbal anchor "Strong" has the approximate position of 50% of "Maximum", and "Moderate" is approximately half of "Srong", etc.

Studies with Psychophysical Scaling and Depression

Our previous studies (Borg, Magalhães, Mörtberg & Costa, 2016 – under review, Magalhães, 2017, Magalhães, Borg, Mörtberg & Costa, 2017) we used the CR100 to assess 32 symptoms of depression. The first study assessed non-depressed Swedish students, the second investigated non-depressed Brazilian students, and the third studied a Brazilian clinical and non-clinical sample.

In all studies the centiMax performed like a very stable scale for assessing depressive symptoms. More than 400 subjects participated in the research (some unpublished yet) and among many advantages of using psychophysical scaling we found: (1) comparisons between and within symptoms, (2) comparisons between and within groups, (3) direct measure of different parts of depressive disorder, (4) the same numbers reflected the same perceptual weight among symptoms, (5) level of difference between symptoms, (6) the ratios between symptoms for different populations (ie. how strong a symptom is for depressive and non-depressive people), (7) possibility of comparison of the raw scores without loss of mathematical information, (8) creation of a symptom profile (Fig 1.), where a graphic showing the "shape" of the depressive disorder can be generated for each subject or group. This allows the professional to better understand which symptoms or groups of symptoms deserves more attention either for psychiatric or psychological intervention. The last study also revealed good psychometric properties that makes this instrument a good possibility of a new questionnaire to assess depression.

All these features appear as a great improvement in psychological assessment if compared with the currently used instruments with Likert-like scales. The way the centiMax was created as a level anchored ratio scale, better reflects the real life experiences for the individuals. For the future, follow up studies must show that the scale is also stable in measuring changes along time, especially for patients under treatment.

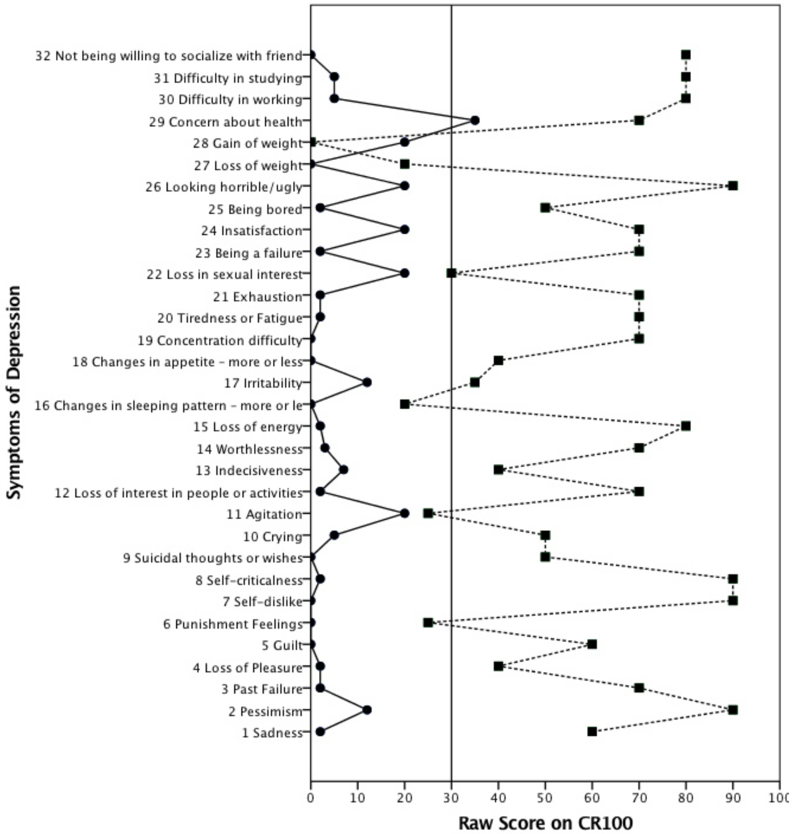


Figure 1. Symptom profiles of two subjects using the raw score of CR100. The full line represents a subject with a mean of 25.4 cM, and the dashed line a subject with a mean of 77.0 cM. The vertical line corresponds to the criteria for moderate depression (30 cM) according to Magalhães (2017).

Final considerations

Because they are reliable methods, psychophysical scaling can be used to monitor the evolution of diseases and the effects of therapeutic treatment. In recent years, Psychophysics has gained more and more space within clinical activities and certainly presents a very promising future as a psychological activity (Costa, 2011). We hope that new areas in Psychology and Psychiatry fields can discover and use psychophysical methods to access mental events and behaviors producing a better and more precise measurement.

References

Association, A. P. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Arlington: American Psychiatric Publishing.

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.* (Doctor degree), Stockholms universitet, Stockholm.

Borg, G. (1962). *Physical performance and perceived exertion*. Lund: Gleerup.

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

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, Magalhães, Mörtberg & Costa, 2016 – under review,

Costa, M. F. (2011). A clínica da psicofisica. *Psicologia USP, 22*(1), 14-44.

Ferris, T. L. J. (2004). A new definition of measurement, *Measurement*, 36(1): 101–9.

Hadžibajramović, E., Svensson, E., & Ahlborg Jr., G. (2013). *Construction of a global score from multi-item questionnaires in epidemiological studies*. Retrieved from Örebro, Sweden.

Hamilton, M. (1960). A Rating Scale for Depression. *Journal of Neurology, Neurosurgery & Psychiatry, 23*, 56-62.

Kornbrot, D. E., Msetfi, R. M., & Grimwood, M. J. (2013). Time Perception and Depressive Realism: Judgment Type, Psychophysical Functions and Bias. *8, 8*, 1-9.

Luce, R. D., & Krumhansl, C. L. (1988). Measurement, scaling, and psychophysics. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey & R. D. Luce (Eds.), *Stevens' Handbook of Experimental Psychology* (2 ed., Vol. 1, pp. 3-74). New York: John Wiley & Sons.

Magalhães, A. (2017). *Três estudos usando a Escala centiMax de Borg® (Escala CR de Borg®, CR 100, cM) para o escalonamento de sintomas depressivos.* (PhD Thesis), Univesity of São Paulo, São Paulo, Brazil.

Magalhães, A.; Borg, E.; Mörtberg, E.; Costa, M. (2017). The use of The Borg CR 100® Scale to assess depressive symptoms in Brazilian and Swedish samples. In: 20th Euro Congress on Psychiatrists and Psychologists. Annals of 20th Euro Congress on Psychiatrists and Psychologists. Rome, Italy: iMedPub LLC.

Montgomery, S. A., Åsberg M. (1979). "A new depression scale designed to be sensitive to change". *British Journal of Psychiatry*. 134 (4): 382–89

Schwenk T. L., Coyne, J. C., Fechner-Bates, S. (1996). Differences between detected and undetected patients in primary care and depressed psychiatric patients. *Gen Hosp Psychiatry*. 18(6):407-15.

Stevens, S. S. (1958a). Measurement and Man. *Science, 127*, 383-389.

Stevens, S. S. (1958b). Problems and Methods of Psychophysics. *Psychological Bulletin, 55*(4), 177-196Stevens, S. S. (1955). On the Averaging of Data. *Science, 121*, 113-116.

Stevens, S. S. (1959). The Quantification of Sensation. *Daedalus, 88*(4), 606-621.

Stevens, S. S. (1975). *Psychophysics*: Introduction to its perceptual, neural, and social prospects. Oxford, England: John Wiley & Sons.

World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. Geneva: World Health Organization.

VALIDITY EVIDENCE OF THE SPECIFIC PICTORIAL SCALE OF PERCEIVED EXERTION FOR FOOTBALL PLAYERS - GOL SCALE: THE PICTORIAL SCALE TO MEASURE THE PERCEIVED EXERTION IN SPORT

Luis Felipe Tubagi Polito[1,2], Elisabet Borg[3], Aylton José Figueira Junior[1], Marcelo Callegari Zanetti[1], Carla Giuliano de Sá Pinto Montenegro[4], Helton Magalhães Dias[1], Douglas Popp Marin[2], and Maria Regina Ferreira Brandão[1]

[1]*Post-Graduate Programme in Physical Education, São Judas Tadeu University, São Paulo, SP, Brazil*

[2]*Medical Sciences and Health School, Methodist University of São Paulo, São Bernardo do Campo, SP, Brazil*

[3]*Department of Psychology, Stockholm University, Stockholm, Sweden*

[4]*Israeli Institute of Teaching and Research Albert Einstein, São Paulo, SP, Brazil*

luis.polito@uol.com.br

Abstract

Despite the methods usually used for ratings of perceived exertion to estimate the internal training load and monitoring the exercise, few instruments have been developed and validated for specific sports using just pictorial anchors. This study aimed: (1) to suggest one specific theoretical model to elaborate a pictorial scale; (2) to analyze the evidences of validation of a scale for football players named GOL Scale. In the first phase of the development the cartoons of the scale were drawn based on the theoretical construction of perceived exertion. The final version of the scale is composed by six cartoons which show different grades of effort. To ensure the representativeness of the cartoons they were validated by nine judges from seven different research lines in Sports Science. In the second phase, thirteen male football players were evaluated by a three-minute progressive protocol up to voluntary exhaustion with a one-minute recovery between stages (Maximal Cardiopulmonary Effort Test – MCET), and by Yo-Yo Intermittent Recovery Test – Level 1. The Borg RPE Scale (6-20), The Cavasini Scale, The cartoon GOL Scale, Heart Rate (HR), Percentage of Heart Rate (%HR) and Blood Lactate Concentration ([La]) were immediately determined after each stage of both tests. The level of significance was 0.05. In the MCET, significant correlation between the GOL Scale and RPE 6 – 20 Borg Scale (r = 0.92), Cavasini Scale (r = 0.93), %HR (r = 0.86), HR (r = 0.80), and blood lactate (r = 0.58) were found, whereas during Yo-Yo Test, significant correlation between the GOL Scale and Borg RPE Scale (r = 0.93), Cavasini Scale (r = 0.93), %HR (r = 0.82), HR (r = 0.80), and Blood Lactate (r = 0.77) were found. These results showed that the GOL Scale is a promissory instrument to measure perceived exertion, without the translation problems commonly found in the verbal scales.

There are many important variables to evaluate the athletic performance during training sessions and matches, such as time, training method, volume, recovery, density, intensity and others. So, the use of specific instruments is necessary to evaluate each one of these characteristics. This way, the rating of perceived exertion (RPE) is a practical and low-cost method which can be used in different sports to monitor the intensity of efforts made on training sessions, such as to estimate the internal training load by the Foster *et al.* (2001) protocol (Impellizeri, Rampinini, and Marcora, 2005).

In the literature there are some instruments which were developed and validated to be applied in different exercises, such as the Borg centiMax Scale® (CR100) (Borg and Borg, 2001; Borg, 2007), and the Borg RPE Scale (6 – 20) (Borg, 1970, 1982), which has fifteen numerical anchors and nine verbal anchors. The Borg RPE Scale is the most common scale used to measure the intensity of the efforts in football, for instance (Polito *et al.*, 2017).

In football and other sports there are different methods to control the training parameters, including more expensive devices, such as gadgets which measure the Global Position System (GPS) or heart rate monitors. According to Polito *et al.* (2017), methods of rating the perceived exertion are practical and low-cost, which make these instruments able to be applied in clubs with different economic conditions around the world. The use of Borg RPE Scale around the world required the translation of this instrument into more than fifteen different languages.

However, some simple translation of the original verbal anchors of the Borg RPE Scale (6 – 20) were also found, which had been done without cross-cultural validation (an important method which allows to keep the scientific validation of the scale). On the other hand, it is very important to take proper care of the scale so that one does not adapt it in a way which can result in its modification (Hambleton, 2005; Borsa, Damásio, and Bandeira, 2012).

The adaptation of psychological or psychophysiological instruments is a complex task which requires careful planning regarding its psychometric properties, keeping their content and validity (Cassepp-Borges, Balbinotti and Teodoro, 2010). An evaluation of the semantic concordance in different cultures is necessary. Nevertheless, the cross-cultural adaptation has not always been executed in translation of the Borg RPE Scale (6 – 20), which can have contributed with the incorrect interpretation and the incorrect use of these instruments.

According to Bar-Or (1989), Robertson *et al.* (2000), Utter, Kang, Nieman, Dumke, and Mcanulty (2006) and Costa *et al.* (2008) one alternative way to deal with the semantic problems of translation of the verbal scales is the elaboration and validation of pictorial scales, without any verbal anchors.

In spite of the fact that the perceptions of exercise-related symptoms are very similar in different sports and exercise, The Scientific Committee of European Society of Sports Psychology (Tenenbaum and Bar-Eli, 1995) has suggested that the psychometric instruments should reflect the single context of the sport, making the ecologic validation of the instrument better. These data are reinforced by Mcguigan (2017).

Therefore, the purpose of this research was: (1) to suggest one specific theoretical model to elaborate a pictorial scale; (2) to analyze the preliminary evidences of validation of a scale for football players named GOL Scale as a reference to the main goal of football, the GOAL (GOL in Portuguese).

Method

The method was divided in two different phases. The first phase was the construction of the GOL Scale, and the second phase was the evaluation of its construction and concurrent validation.

In order to construct the GOL Scale, the most common acute physiological signs during effort were taken into account, including: ventilation modification provided by the elevation in expired carbon dioxide; facial flushing generated by the thermoregulation process; and possible biomechanical and postural changes provided by different fatigue levels. These signs were elaborated on six different images, two images for each exercise phase according to the multiple anaerobic threshold model proposed by Skinner and McLellan (1980).

Using these first parameters, the first sample of the GOL Scale was constructed by a professional cartoonist (version A of figure 1).

However, in order to minimize the subjective influence in the design of the instrument, this scale was sent to nine PhD professors (judges) from seven different research lines in Sports Science: sports training, cardiovascular physiology, football physiology, sports psychology, biomechanics, metabolism and exercise psychophysiology. These professors were asked to suggest the addition or exclusion of one or more images, or just the modifications in the existing images. After the first evaluation, further details in ventilation and facial flushing were required, which resulted in a second version of the GOL Scale (version B of figure 1).

In the second evaluation the biomechanics specialist called attention to the knee extension, which is uncommon in football abilities. In addition to this, the authors suggested the inclusion of numerical anchors in the scale for the third version (version C of figure 1).

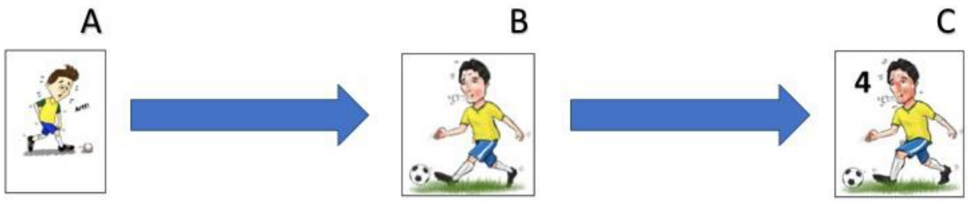


Fig. 1 The development of the fourth cartoon of the GOL Scale.

The last recommendation of the author was the change of the uniform color, reducing the “Brazilian” characteristics in it, giving the scale a more international design.

In the second phase thirteen male football players (18.8 ± 0.77 years old, stature 177 ± 8.0 cm, body mass 70.8 ± 7.53 kg, % body fat 13.42 ± 3.19, lean mass 60.49 ± 4.75 kg and fat mass 9.60 ± 3.19 kg) were evaluated by a three-minute progressive protocol up to voluntary exhaustion with one-minute recovery between stages (Maximal Cardiopulmonary Effort Test – MCET), and by Yo-Yo Intermittent Recovery Test – Level 1. The Borg RPE Scale (6 – 20), The Cavasini Scale, The cartoon GOL Scale, Heart Rate (HR), Percentage of Heart Rate (%HR) and Blood Lactate Concentration ([La]) were immediately determined after each stage of both tests. Pearson's correlations coefficients were used. The level of significance was 0.05.

Results and Discussion

The association with objective parameters such as heart rate, blood lactate concentration and oxygen consumption is frequently employed as a method in validation of different perceived exertion scales (Chen, Fan and Moe, 2002). It is also important to evaluate the association between the GOL Scale and other instruments validated to this purpose, such as the Borg RPE Scale (6 – 20).

In the MCET, significant correlation between the GOL Scale and RPE 6 – 20 Borg RPE Scale (6 – 20) (r = 0.92), Cavasini Scale (r = 0.93), %HR (r = 0.86), HR (r = 0.80), and blood lactate (r = 0.58) were found, whereas during Yo-Yo Test, significant correlation between the GOL Scale and Borg RPE Scale (6 – 20) (r = 0.93), Cavasini Scale (r = 0.93), %HR (r = 0.82), HR (r = 0.80), and Blood Lactate (r = 0.77) were found. The results for each anchor of the GOL Scale in both tests can be found in tables 1 and 2.

The association method used to validate GOL Scale has previously been used for other pictorial scales and regarded as sufficient for validation (Yelling, Lamb and Swaine, 2002; Pfeiffer, Pivarnik, Womack, Reeves and Malina, 2002; Robertson *et al.*, 2003; Robertson *et al.*, 2004; Robertson *et al.*, 2005; Micklewright, Gibson, Gladwell and Al Salman., 2017).

Therefore, the strong correlation values identified from the association of GOL Scale with objective parameters of effort intensity, as well as the association with the other validated constructs used to evaluate the perceived exertion provide preliminary data of the validation of the GOL Scale for football athletes.

Conclusion

It is suggested that the absence of the verbal anchors contribute to a broader application of the GOL Scale for football athletes. At the same time, future researches should evaluate the application of this instrument for athletes with different languages and literacy levels, confirming or refuting the initial hypothesis of this research.

Table 1. Physiological and perceptual variables in each phase of GOL Scale during the MCET

GOL Scale	HR (bpm)	$\%HR_{máx}$	VO_2 (ml/kg/min)	$\%VO_{2máx}$	[La]	BORG	CAVASINI
1	120 ± 23	63,65 ± 15,78	20,96 ± 5,99	46,28 ± 11,80	3,42 ± 1,42	7 ± 1	1 ± 1
2	142 ± 18	74,84 ± 9,06	28,31 ± 6,31	62,89 ±9,22	3,33 ± 1,29	9 ± 2	3 ± 1
3	163 ± 18	85,14 ± 7,49	33,21 ± 5,67	77,97 ± 8,68	4,69 ± 2,93	11 ± 2	4 ± 1
4	174 ± 17	89,45 ± 6,45	36,54 ± 6,04	83,22 ± 8,69	4,82 ± 2,06	13 ± 2	5 ± 1
5	179 ± 15	94,65 ± 4,41	39,28 ± 7,63	89,49 ± 10,12	6,79 ± 3,82	15 ± 1	7 ± 1
6	188 ± 14	99,16 ± 1,75	40,23 ± 6,70	97,07 ± 5,47	8,24 ± 3,88	18 ± 1	9 ± 1

Table 2. Physiological and perceptual variables in each phase of GOL Scale during the Yo-Yo Intermittent Recovery Test – Level 1

GOL Scale	HR (bpm)	%$HR_{máx}$	VO_2 (ml/kg/min)	%$VO_{2máx}$	[La]	BORG	CAVASINI
1	132 ± 21	71,47 ± 11,97	37,52 ± 1,20	73,72 ± 8,69	4,17 ± 5,04	7 ± 1	1 ± 1
2	151 ± 13	81,37 ± 7,98	39,66 ± 2,27	74,35 ± 9,07	4,40 ± 1,78	8 ± 2	2 ± 2
3	168 ± 10	90,56 ± 5,13	43,74 ± 4,76	80,08 ± 7,80	5,96 ± 2,43	10 ± 1	4 ± 1
4	176 ± 12	93,93 ± 3,77	45,02 ± 4,70	85,14 ± 5,79	7,23 ± 2,00	12 ± 2	6 ± 2
5	183 ± 11	95,81 ± 3,74	48,16 ± 5,06	90,75 ± 4,34	9,41 ± 3,25	16 ± 2	7 ± 1
6	184 ± 12	98,80 ± 2,37	53,72 ± 7,66	98,00 ± 3,17	10,61 ± 2,76	19 ± 2	9 ± 1

Acknowledgements

We would like to thank São Judas Tadeu University (USJT), the Coordination of Improvement of Higher Education Personnel (CAPES), Stockholm University (SU) and Methodist University of São Paulo (UMESP) for their collaboration with this research. In addition, we would like to thank the organizing committee of the Fechner Day for the stipends which made this presentation possible.

References

Bar-Or, O., & Ward, D. S. (1989). Rating of perceived exertion in children. In Bar-Or, (Org.) Advances in pediatric sport sciences (pp. 151-168), Champaign.

Borg, G. (1970). Perceived exertion as an indicator of somatic stress. Scand j rehabil med, 2, 92-98.

Borg, G. A. V. (1982). A category scale with ratio properties for intermodal and interindividual comparisons. Psychophysical judgment and the process of perception, 25-34.

Borg, G., & Borg, E. (2001). A new generation of scaling methods: Level-anchored ratio scaling. Psychologica, 28(1), 15-45.

Borg, E. (2007). On perceived exertion and its measurement (Doctoral dissertation, Psykologiska institutionen).

Borsa, J. C., Damásio, B. F., & Bandeira, D. R. (2012). Cross-cultural adaptation and validation of psychological instruments: some considerations. Paidéia (Ribeirão Preto), 22(53), 423-432.

Cassepp-Borges, V., Balbinotti, M. A. A., & Teodoro, M. L. M. (2010). Tradução e validação de conteúdo: Uma proposta para a adaptação de instrumentos. In L. Pasquali, Instrumentação psicológica: Fundamentos e práticas (pp. 506-520). Porto Alegre: Artmed.

Chen, M. J., Fan, X., & Moe, S. T. (2002). Criterion-related validity of the Borg ratings of perceived exertion scale in healthy individuals: a meta-analysis. Journal of sports sciences, 20(11), 873-899.

Costa, M. G., Dantas, E. H. M., Marques, M. B., & da Silva Novaes, J. (2008). Subjective exertion perception. Perceived exertion classification: face scale utilization proposal. Fitness & Performance Journal, 3, 305-313.

Foster, C., Florhaug, J. A., Franklin, J., Gottschall, L., Hrovatin, L. A., Parker, S., Doleshal, P., & Dodge, C. (2001). A new approach to monitoring exercise training. The Journal of Strength & Conditioning Research, 15(1), 109-115.

Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), Adapting educational and psychological tests for cross-cultural assessment (pp. 3-38). Mahwah, NJ: Lawrence Erlbaum.

Impellizzeri, F. M., Rampinini, E., & Marcora, S. M. (2005). Physiological assessment of aerobic training in soccer. Journal of sports sciences, 23(6), 583-592.

McGuigan, M. (2017). Monitoring training and performance in athletes. Human Kinetics.

Micklewright, D., Gibson, A. S. C., Gladwell, V., & Al Salman, A. (2017). Development and Validity of the Rating-of-Fatigue Scale. Sports Medicine, 47(11), 2375-2393.

Pfeiffer, K. A., Pivarnik, J. M., Womack, C. J., Reeves, M. J., & Malina, R. M. (2002). Reliability and validity of the Borg and OMNI rating of perceived exertion scales in adolescent girls. Medicine & Science in Sports & Exercise, 34(12), 2057-2061.

Polito, L. F. T., Figueira Jr, A. J., Miranda, M. L. J., Chtourou, H., Miranda, J. M., & Brandão, M. R. F. (2017). Psychophysiological indicators of fatigue in soccer players: A systematic review. Science & Sports, 32(1), 1-13.

Robertson, R. J., Goss, F. L., Boer, N. F., Peoples, J. A., Foreman, A. J., Dabayebeh, I. M., ... & Thompkins, T. A. Y. L. O. R. (2000). Children's OMNI scale of perceived exertion: mixed gender and race validation. Medicine and science in sports and exercise, 32(2), 452-458.

Robertson, R. J., Goss, F. L., Rutkowski, J., Lenz, B., Dixon, C., Timmer, J., Frazee, K., Dube, J., & Andreacci, J. (2003). Concurrent validation of the OMNI perceived exertion scale for resistance exercise. Medicine & Science in Sports & Exercise, 35(2), 333-341.

Robertson, R. J., Goss, F. L., Dube, J., Rutkowski, J., Dupain, M., Brennan, C., & Andreacci, J. (2004). Validation of the adult OMNI scale of perceived exertion for cycle ergometer exercise. Medicine & Science in Sports & Exercise, 36(1), 102-108.

Robertson, R. J., Goss, F. L., Andreacci, J. L., Dubé, J. J., Rutkowski, J. J., Snee, B. M., Kowallis, R. A., Crawford, K., Aaron, D. J., & Metz, K. F. (2005). Validation of the children's OMNI RPE scale for stepping exercise. Medicine & Science in Sports & Exercise, 37(2), 290-298.

Skinner, J. S., & McLellan, T. H. (1980). The transition from aerobic to anaerobic metabolism. Research quarterly for exercise and sport, 51(1), 234-248

Tenenbaum, G., & Bar-Eli, M. (1995). Contemporary issues in exercise and sport psychology research. European perspectives on exercise and sport psychology., 292-323.

Utter, A. C., Kang, J., Nieman, D. C., Dumke, C. L., & Mcanulty, S. R. (2006). Validation of Omni scale of perceived exertion during prolonged cycling. Medicine and science in sports and exercise, 38(4), 780-786.

Yelling, M., Lamb, K. L., & Swaine, I. L. (2002). Validity of a pictorial perceived exertion scale for effort estimation and effort production during stepping exercise in adolescent children. European Physical Education Review, 8(2), 157-175.

SESSION I

HEMISPHERIC EVOKED POTENTIALS AND SUBJECTIVE THRESHOLD OF AUDITORY STIMULI FOR HUMAN

Kazi Saifuddin
Department of Psychology, Jagannath University, Dhaka, Bangladesh
Nazmul A. Khan
Department of Environmental Science and Management, North South University, Dhaka, Bangladesh
kazisaifuddin@ymail.com

White noises as auditory stimuli were tested psychologically and physiologically for human subject. When the white noise stimuli were tested psychologically the two factors of autocorrelation function (ACF) were as parameter. The ACF was defined here, as $\emptyset(\tau) = \frac{1}{2T}\int_{-T}^{+T} p(t)p(t+\tau)dt$, where, $p(t)$ is the signal (stimulus) at the entrance of the ears is the delay time, and $2T$ is the integration interval. In such ACF analysis of acoustic signal, four parameters can be found, namely:

a. the energy represented at the origin of the delay, $\Phi(0)$;
b. the amplitude, $\phi 1$, between the first peak and the zero crossing number;
c. the structure including the time delay of the first peak; and
d. the effective duration of the envelope of the normalized ACF, τe, which is defined by ten percentile delay or at which the envelop of the ACF becomes -10dB.

The first (a) and second (b) parameters were tested psychologically in anechoic chamber and found significant variations only for first parameter (a) in the perception of auditory time duration as 50% threshold levels in the psychometric functions, but not for second parameter (b). The psychometric functions were extracted by using cumulative presentation of the data of subjective paired-comparison test on auditory temporal durations.

Hemispheric dominance was also investigated simultaneously for auditory temporal perception as physiological test of human subject. In the anechoic chamber auditory evoked potentials (AEP) for the human subject were recorded by using conventional EEG and by the summing computer techniques to study the effects of above two parameters (a, and b) of ACF. Thus, auditory stimuli (white noise) were propagated with two different factors maintaining convention of anechoic chamber. Results revealed that dominating amplitudes of AEP wave forms in average were found in respect to the first parameter (a) of ACF for the left-hemisphere than that of the right-hemisphere. No visible difference was found for the second parameter (b) of ACF. Also 50% threshold differences were found for parameter ‘a’ in amplitude but not in latency. The dominating latencies of AEP were found for the left-hemisphere than that of the right-one also for parameter ‘a’. But not any effects of parameter ‘b’ found in physiological test. Therefore, the left-hemisphere was concluded as dominating than that of the right one for the auditory-temporal perception physiologically and also related to that of the subjective threshold for auditory time duration psychologically.

Keywords: Auditory temporal duration, auditory evoked potential, autocorrelation function

MAJOR AND MINOR CHORDS PERCEPTION BY PROFESSIONAL MUSICIANS AND NON-MUSICIANS

Stanislav O. Skorik and Nickolay A. Almayev
Institute of Psychology, Russian Academy of Sciences, Moscow, Russian Federation
stanskorik@mail.ru, almaev@mail.ru

Abstract

Previously we found that attenuation pattern of a signal (amplitude envelope) plays significant role in perception of isolated major and minor chords. Four types of modified chords – without attenuation, linear attenuation, MIDI, and real (piano) were administered to the two groups of subjects – non-musicians (N=29) and professional musicians (N=14). Both groups estimated stimuli by 35 subjective unipolar scales. Non-musicians were mostly unable to identify chords other than real, while professionals did it successfully. Factor analysis has shown structure of non-musician's responses consisting of three factors: 1) demand for resources to cope with sound, 2) activation, cheering up, and 3) minor. Factor structure of the professional's results has three factors: 1) major-minor, 2) demand for resources, 3) affiliative activation by minor. In both groups subjectively identified minor was connected to scales "help it", "pleading".

Musical acoustic events perfectly convey emotional states, that was repeatedly shown in many of scientific experimental studies. But until now psychological mechanisms of it are not completely clear. One of the basic elements of musical sound most studied in this context is the tonality i.e. major and minor frets of contemporary European music. In music psychology major is associated with positive emotional valence, while minor - with the negative one. Traditionally, such studies are based on the evaluation of musical compositions, their samples, isolated melodies or harmonic sequences. However, studies with factor design have shown that other characteristics of sound, such as tempo and harmony, can influence the valence of musical samples estimates (Farbood, 2012). There are also works in which the assessment of the valence of musical compositions in preschool age was associated with the speed (tempo) of a melody played (Gerardi & Gerken, 1995). Psychological mechanisms of musical stimuli perception may be studied in depth by isolated variation of one, two parameters of them under the control of all of the others (Almayev et al., 2017).

Triad chords are the basis of the harmonic component of music, they already have those pitches, basing on which the subjects can determine major and minor frets and musical scales.

The ability of subjects without musical education and the experience of playing on musical instruments to differentiate triad chords into happy and sad ones is usually not questioned. In the scientific literature, even can be found such studies in which people with amusia cope with this task (Marin et al., 2015). But at the same time, it is not completely clear how subjects who do not know the music theory and who have difficulties in distinguishing a triad chord (3 simultaneous notes) from a separate note can quite accurately differentiate the chords using the valence feature.

In our previous work it was shown that one of these mechanisms is related to attenuation of the chord (Skorik, Almayev & Bessonova, 2017). Subjects coped with the task of differentiation only in those groups of stimuli, where some attenuation in the triad chords was presented. However, it is quite possible that our stimuli could not be differentiated by the subjects with musical education and experience in playing on musical instruments. In this situation, it could be stated that attenuation in the case of triad chords is the critical condition for appearance of emotional evaluations. But if musicians are still able to evaluate isolated triad chords without attenuation, this will indicate existence of the different mechanisms of emotion generation in these two groups of subjects.

Method

Four different types of attenuation were investigated (Figure 1):

a) with rectangle form – no attenuation, signal doesn't change; b) with linear (triangle) form of attenuation; c) - midi synthesized; d) real chords (from soundbanks) taken by professional performers. Total number of stimuli - 32, generated on the base of 4 major and 4 minor chords from A5 to D5, administered all in one balk quasi randomly.

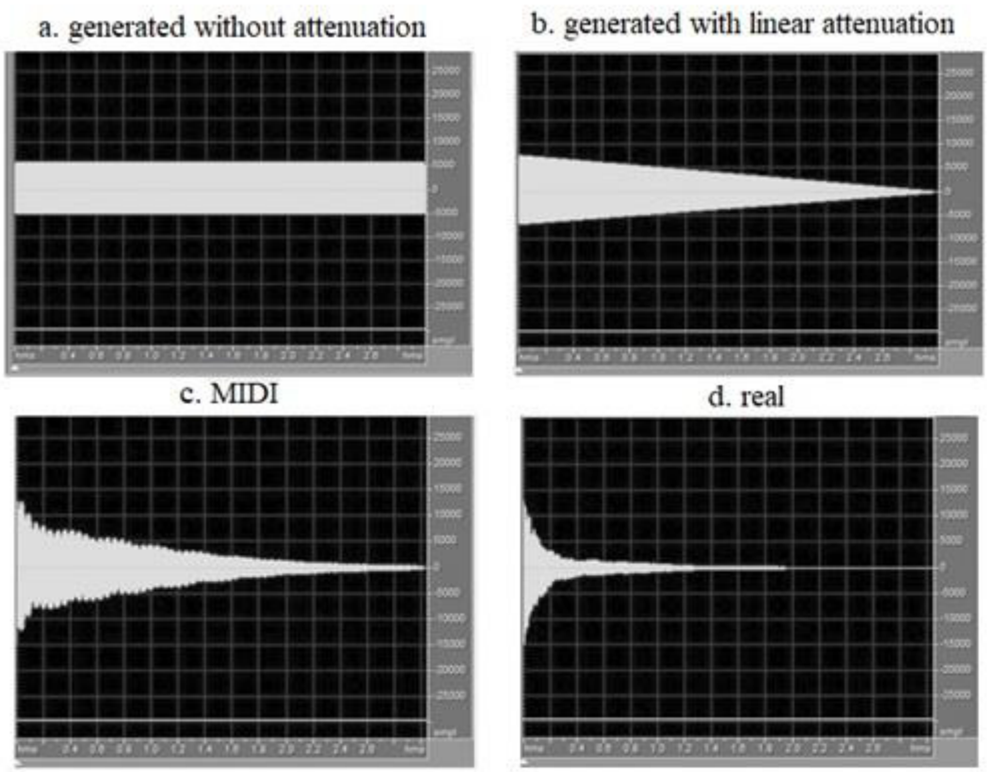


Fig. 1. Stimuli (from left to right, upper row) a. generated without attenuation, b. generated with linear attenuation, (from left to right, lower row) c. MIDI, d. real.

10 people took part in the first part of the study. They evaluated the sounds presented in quasi-random order by the method of free description. 29 subjects (14 women and 15 men; medium age – 26,7) assessed chords in the second part by 35 unipolar scales (Table 1), which were constructed on the basis of data from the first part of the study. It should be noted that subjects mostly had not any musical education and didn't practice music as amateurs. Group of subjects-musicians assessed chords by the same way and consisted of 14 persons (6 women and 8 men; medium age – 28,2). They all were professional musicians with different levels of training and had experience of live performances.

Results and Discussion

Differences in differentiating triad chords according to valence in non-musicians and musicians

The scores of the subjects without musical education and the musicians were compared according to Wilcoxon's criterion (Table 1).

In the group of stimuli without attenuation non-musicians failed to find any differences between major and minor triad chords, while musicians coped with this task by 15 subjective scales.

In the group of generated stimuli with a linear attenuation, non-musicians differentiated major and minor by 4 subjective scales, and the musicians by 9.

In the group of MIDI-stimuli, non-musicians differentiated major and minor by 3 subjective scales, while the musicians by 23.

In the group of real stimuli with linear attenuation, non-musicians differentiated major and minor by 24 subjective scales, and the musicians by 26.

Comparison of these results shows that professional musicians look for the tonal structure and make their judgments, basing on it. Professional musicians differentiate major and minor triad chords, even when the character of signal attenuation is very specific, and hardly was previously even heard by them (linear attenuation in the second group of stimuli). Nevertheless, the type of attenuation helps professional musicians to cope with the task better: 26 scales under the differentiation of real sounds vs 8 scales in the differentiation of sounds with linear attenuation. Non-musicians successfully coped with the task only in the case of the real sounds - and they did it not much worse than professional musicians. Thus, it can be concluded that the specific nature of attenuation characteristic of real signals, facilitates manifestation of the features of the tonal structure of major and minor, although professionals unlike non-musicians, less bond to attenuation patterns.

№	Scale	Non-musicians				Musicians			
		Without attenuation	Linear attenuation	MIDI	Real sounds	Without attenuation	Linear attenuation	MIDI	Real sounds
1	Pleasant								
2	Tension								
3	Want to do something								
4	Heavy								
5	Bright								
6	Active								
7	Anxious								
8	Happy								
9	Sad								
10	Expected that something will happen								
11	Warm								
12	Major								
13	Male / Feminine								
14	Interesting								
15	Muffle it								
16	Natural								
17	Peppy								
18	Optimistic								
19	Energetic								
20	Confident								
21	Pleading								
22	Exciting								
23	Help it								
24	Cold								
25	Minor								
26	Musical								
27	Kind								
28	Disruptive								
29	Unkind								
30	Positive								
31	Boring								
32	Relaxing								
33	Working properly								
34	Loud								
35	Old								

Tab. 1. Unipolar scales, used in study. Shaded and spotted filled cells (non-musicians and musicians, respectively) indicate scales, according to Wilcoxon's criterion ($p<.001$) allow the subjects to differentiate major and minor.

Results of factor analysis

Factor analysis (Varimax raw) for two groups of subjects was carried.

In the group of non-musicians according to the results of the Scree-test, 3 factors were identified, explaining in general 50% of variance.

1. The first factor "General Tension" (30.4% of variance), i.e. demand for resources to cope with sound, consists of: *"pleasant" (.80), "kind" (.78), "positive" (.77), "bright" (.71), "happy" (.68), "relaxing" (.68), "optimistic" (.67), "warm" (.66), "interesting" (.62), "musical" (.62), "anxious" (-.68), "unkind" (-.68), "heavy" (-.71), "tension" (-.73), "muffle it" (-.74).*

2. The second factor "Activation" (13.5% of variance) consists of: *"energetic" (.81), "peppy" (.80), "confident" (.73), "exciting" (.72), "active" (.68), "pleading" (-.21), "minor" (-.22), "old" (-.24), "sad" (-.36), "boring" (-.47).*

3. The third factor “Minor” (6.1% of variance) consists of: *“help it” (.63), “pleading” (.61), “minor” (.57), “confident” (-.28), “working properly” (-.34).*

In the group of musicians according to the results of the Scree-test, 3 factors were identified, explaining in general 50% of variance.

1. The first factor “Extended semantic description of major / minor” (31% of variance) consists of: *“major” (.82), “optimistic” (.81), “happy” (.80), “energetic” (.79), “peppy” (.76), “confident” (.71), “active” (.67), “positive” (.62), “exciting” (.57), “warm” (.55), “bright” (.54), “kind” (.41), “interesting” (.40), “old” (-.43), “pleading” (-.48), “boring” (-.53), “cold” (-.55), “sad” (-.77), “minor” (-.81).*

2. The second factor, “Extended General Tension” (11.3% of variance) consists of: *“pleasant” (.81), “musical” (.69), “kind” (.69), “interesting” (.64), “relaxing” (.63), “working properly” (.60), “natural” (.59), “warm” (.53), “bright” (.48), “positive” (.41), “old” (-.40), “boring” (-.43), “disruptive” (-.54), “anxious” (-.57), “unkind” (-.58), “heavy” (-.63), “muffle it” (-.76), “tension” (-.78).*

3. The third factor “Affiliative activation” (if not major, then “help it”) (6.7% of variance) consists of: *“help it” (.67), “expected that something will happen” (.59), “want to do something” (.59), “pleading” (.55), “exciting” (.51), “anxious” (.44), “boring” (-.11), “old” (-.13), “major” (-.19).*

The main differences between the factor solutions of non-musicians and musicians are in the first factor. For musicians it is the extended semantics of major and minor, for non-musicians - the general tension. For musicians, the factor of general tension becomes the second. And the third factor for non-musicians and musicians has common features: it is a factor of minor (non-major for musicians) with the main scales “pleading” and “help it”. The specificity of the factor solution in the case of professional musicians is that non-majority is related to the activation and expectation scales.

Acknowledgements

Our studies since 2016 were supported by RFBR grant # 16-06-00487.

References

Almayev N.A., Skorik S.O., Medincev A.A., Bessonova Yu.V., Sadov V.A., Murasheva O.V. (2017). *Resource-based Approach to Music Psychology. Proccedings of the 6th Conference of The Asia-Pacific Society for the Cognitive Sciences of Music (APSCOM 6)*, 35-41.

Farbood, M. M. (2012). *A parametric, temporal model of musical tension. Music Perception: An Interdisciplinary Journal,* 29(4), 387-428.

Gerardi, G. M., & Gerken, L. (1995). *The development of affective responses to modality and melodic contour. Music Perception,* 12, 279–290.

Marin, M. M., Thompson, W. F., Gingras, B., & Stewart, L. (2015). *Affective evaluation of simultaneous tone combinations in congenital amusia. Neuropsychologia,* 78, 207-220.

Skorik S.O., Almayev N.A., and Bessonova Yu.V. (2017). *The character of sounds attenuation as the decisive factor in Major and Minor chords perception. Proceedings of The 33rd Annual Meeting of the International Society for Psychophysics Fechner Day 2017*, 22–26.

SEARCHING FOR A MEASURE OF AUDITORY TEMPORAL RESOLUTION: PRELIMINARY DATA

Shuji Mori[1], Ryota Aratake[2], Kazuhito Ito[3], Yoh-ichi Fujisaka[4], Takashi Morimoto[4], Shunsuke Tamura[5], and Nobuyuki Hirose[5]

[1] *Department of Informatics, Faculty of Information Science and Electrical Engineering, Kyushu University*

[2] *Department of Electrical Engineering and Computer Science, School of Engineering, Kyushu University*

[3]*Department of Information Science, Faculty of Liberal Arts, Tohoku Gakuin University*

[4] *RION Co., Ltd.*

[5] *Department of Informatics, Graduate School/Faculty of Information Science and Electrical Engineering, Kyushu University*

mori@inf.kyushu-u.ac.jp

Abstract

This paper reports a preliminary study on auditory temporal resolution that measured within- and across-frequency gap thresholds and amplitude modulation thresholds in a group of young healthy listeners. We analyzed the relations among the obtained measures of temporal resolution. The results revealed that, while two methods of amplitude modulation yielded similar results, there were some discrepancies among the various measures of temporal resolution.

The ability to detect temporal changes in sounds, or auditory temporal resolution, is vital for everyday listening of complex acoustic patterns, such as speech. There is a growing recognition that hidden hearing loss, cochlear neuropathy which is undetectable by assessments of hearing levels, impairs temporal processing of incoming sound and might result in difficulties in understanding speech in noise (Oxenham, 2016; Plack, Barker, & Prendergast, 2014). The standard audiometry test battery used in Japan does not include the measurement of temporal resolution, which is needed for the diagnosis of patients who might suffer from hidden hearing loss or other disorders of temporal processing.

Psychophysically, auditory temporal resolution is mainly measured by two methods, amplitude modulation (AM) detection and gap detection. In the typical AM detection method, the signals are sinusoidally modulated on a noise carrier, and the minimum detectable depth of modulation, or AM threshold, is measured and used to construct temporal modulation transfer function (TMTF), which relates the AM threshold to the modulation frequency (f_m). TMTF is approximated by a first-order low-pass filter and modelled by a function:

$$\emptyset(f_m) = S_p - 10\log_{10}\{1/[1 + (f_m/f_c)^2]\} \qquad (1),$$

where S_p and f_c are the peak sensitivity and cutoff frequency, respectively. f_c can be converted to a time constant (τ), which represents the temporal resolution of the auditory system, by

$$\tau = 1/2\pi f_c \qquad (2)$$

(Eddins & Green, 1995). Eddins & Green (1995) summarized the estimates of τ reported in the literature, which ranged from 2 to 3 msec.

One obstacle to the clinical use of TMTF is multiple measurements of AM thresholds that are required to construct a low-pass filter-like TMTF: 5 or more measurements over a frequency range that can span more than 1000 Hz, which takes half an hour or longer. This is time-consuming and undesirable in the clinical setting. Alternative procedures, involving fewer measurements, have been proposed (Morimoto et al., 2018; Shen & Richards, 2013). In a recent study, Morimoto et al. (2018) proposed a two-point method, which was to measure an AM threshold at a modulation frequency of 8 Hz and a just noticeable modulation frequency at a fixed modulation depth and obtain the best-fit low-pass filter to these two measures (see the Method section for details). Morimoto et al. (2018) reported that the two-point method only took about 10 min and yielded estimates of the two filter parameters, which were similar to and significantly correlated with those obtained from the multiple measurements of AM thresholds for the same listeners.

In the gap detection method, a silent gap is inserted into an otherwise continuous sound, and the minimum detectable gap duration, or gap threshold, is measured. The gap threshold is usually found to be within 5 msec, supporting the idea that the gap threshold also represents τ, the time constant of the auditory system (Eddins & Green, 1995; Forrest & Green, 1987). However, discrepant findings have been obtained when both the AM and gap detection methods are employed for same listeners. Formby & Muir (1988) found no significant correlation between the mean gap thresholds and the estimates of τ from f_c in 6 listeners but they obtained a significant negative correlation between the mean gap thresholds and the mean S_p of TMTF. Shen & Richard (2013) reported similar results for separate groups of 8 to 19 listeners. Shen (2014) found no significant correlation between the gap thresholds and f_c or S_p in 9 young listeners, 24 aged listeners, or both groups combined.

Another complication is that the listener's performance in the AM and gap detection methods is dependent on the acoustic properties of the signals used. Most notable is a frequency difference between the leading and trailing markers in gap detection. Such 'across-frequency' gap thresholds sometimes reach up to 50 msec when the frequency difference is over 2 octaves (Ito et al., 2016; Mori et al., 2015; Phillips et al., 1997). Previous studies have shown significant correlations between the across-frequency and within-frequency gap thresholds for large groups of listeners (Mori et al., 2015; Phillips & Smith, 2004). No attempt has been made to examine the relation between across-frequency gap detection and AM detection.

This was a preliminary study in an ongoing research project, led by the first author and whose ultimate goal is to establish a measure of auditory temporal resolution that can be used for clinical assessment of impaired temporal processing and related disorders. Achieving this goal requires several issues to be addressed, i.e., an appropriate balance between measurement reliability and efficiency, clarification of relations among different measures, finding of robust indices of impaired temporal resolution, etc. This study was not intended to address all of these issues. Rather, it was an exploratory attempt to find out 'what we should do from here'. We conducted two AM detection methods, a standard, multiple-measurement method and a two-point method proposed by Morimoto et al. (2018), and one within- and two across-frequency gap detection methods, in a relatively large number of young listeners.

Method

Participants

Participants were 25 students, 6 females and 19 males, with a mean age of 21.1 years ranging from 18 to 26. Their pure-tone audiometric thresholds were lower than 25 dB between 250 and 8000 Hz in the tested (left) ears, and their mean pure-tone average (PTA) for 500, 1000, 2000, and 4000 was 6.0 dB in a range between 1.25 and 15. None of them reported difficulty hearing any of the stimuli used in the experiment. Written informed consent was obtained from all participants prior to their participation. This study was approved by the Research Ethics Board of the Faculty of Information Science and Electrical Engineering, Kyushu University.

Apparatus and stimuli

Stimuli were generated by a personal computer (HP, ProBook 4540s) using MATLAB software, which also controlled stimulus presentation and data collection. All of the stimuli were presented through an attenuator (Tucker Davis Technologies, PA5) to the listener's left ear at 60 dB SPL via headphones (SENNHEISER, HD 380PRO). Prior to each experimental session, sound pressure levels were calibrated by playing out individual signals to be used and recording their levels using a Brüel & Kjaer sound level meter (model 2250) with a 1/2-inch condenser microphone (model 4192) placed inside an artificial ear (model 4153), which was used as an acoustic coupler.

For the AM detection, stimuli were constructed on a broadband noise carrier (20 to 14,000Hz), with a duration of 500 msec (cos rise/fall 2.5 msec). f_m was set to either 8, 16, 32, 64, 128, 256, or 512 Hz, with the modulation depth (m) manipulated. For the within-frequency gap detection, both the leading and trailing markers of a silent gap were broadband noise (same as the AM carrier). For the across-frequency gap detection, both markers were 0.5-octave band-passed noises, with center frequencies of 1000 Hz and 4000 Hz, respectively. Each marker had a duration of 250 msec (cos^2 rise/fall 3 msec).

Procedure

The experiment was conducted in a sound-attenuated room. For the AM detection, the standard and two-point methods were employed. In the standard method, the AM thresholds were measured using a 3-inteval forced choice (3IFC), 2-up 1-down procedure, in which a modulated signal was presented at either one of the three intervals (each separated by a 500-ms silent period), and an unmodulated signal was presented at the other two intervals. The listener's task was to indicate the interval containing the modulated signal. The modulation depth (m, expressed in dB, $20\log_{10}(m)$) was initially set at 0 dB and changed by 4 dB for the first 4 reversals and 2 dB for the remaining reversals. A single measurement continued until 12 reversals, and the average of the last 8 reversals was taken as the threshold for that session. Measurements were conducted at seven f_m (from 8 to 512 Hz in that order), and TMTF was constructed from the measured thresholds and fitted by Eq. (1) to estimate S_p and f_c. In the two-point method, the AM threshold was first measured at f_m of 8 Hz, using the same procedure as was employed in the standard method. Next, the just noticeable modulation frequency (Δf_m) was measured at a fixed value of m, which was equivalent to one-half of the 8-Hz AM threshold (in dB), using 3IFC, 2-up 1-down procedure. f_m was initially set at 8 Hz with a step size of 2 octaves for the first 4 reversals and of 1 octave employed thereafter. The average of the last 8 reversals was taken as Δf_m. Using the 8-Hz AM threshold as S_p (termed S_{p_2P}) in Eq. (1), we calculated the corresponding cutoff frequency f_{c_2P} by

$$f_{c\_2P} = \Delta f_m \left(10^{-S_{p\_2P}/20} - 1\right)^{-1/2} \quad (3).$$

For the derivation of Eq. (3), see Morimoto et al. (2018).

For both the within- and across-frequency gap detection, the threshold was measured using a 3IFC, 1-up 2-down procedure, in which two markers with a silent gap between them were presented at either one of the three intervals (each separated by a 500-msec silent period), and markers without a gap were presented at the other two intervals. The listener's task was to indicate the interval containing the signal with a gap. The initial gap size was set differently depending on the listeners and stimulus parameters, and the step size was 2 msec for first 4 reversals and 1 msec thereafter. The average of the last 8 reversals (in a total of 12) was taken as the gap threshold.

For both the AM and gap detection, one measurement was made for each set of stimulus parameters. Since the experiment consisted of two AM detection methods (the standard and the two-point methods) and three gap detection methods (one within-frequency and two across-frequency), they were employed in the order of gap, AM, gap, AM, gap, with the order of the two AM and three gap detection randomized separately for each listener. Before the experiment, the listeners were given various amounts of practice[1].

Results

On average, the standard AM detection method took 29.3 min (from 22 to 44 min) to complete, while the two-point method took 7.6 min (from 6 to 11 min). Figure 1 presents (a) the mean S_p obtained with those two methods and (b) time constants τ, calculated from f_c based on Eq. (2), and the gap thresholds. The two mean S_p were not significantly different from each other, $t(24) = 1.95$, $p = 0.06$. A repeated-measure one-way ANOVA of the time constants and gap thresholds revealed a significant main effect, $F(4, 96) = 109.96$, $p = 0.001$, after Huynh-Feldt correction for a lack of sphericity. Multiple comparisons with an overall α level of 0.05 being preserved by Bonferroni correction showed that τ was not significantly different between the two AM detection methods; the within-frequency gap threshold was significantly different from τ of the standard method but not from that of the two-point method; and the two across-frequency gap thresholds were not significantly different from each other but different from all of the other measures.

[1] Seventeen of 25 listeners underwent approximately 4 min of practice of 8-Hz AM detection, while the remaining 8 listeners were allowed full practice sessions of 8- and 512-Hz AM detection. We compared their performance in the two AM detection methods and found no significant difference in f_c. The two groups were significantly different in S_p, but the difference was not significantly interacted with the two methods. Thus, we decided to combine their data into a single dataset for subsequent analyses.

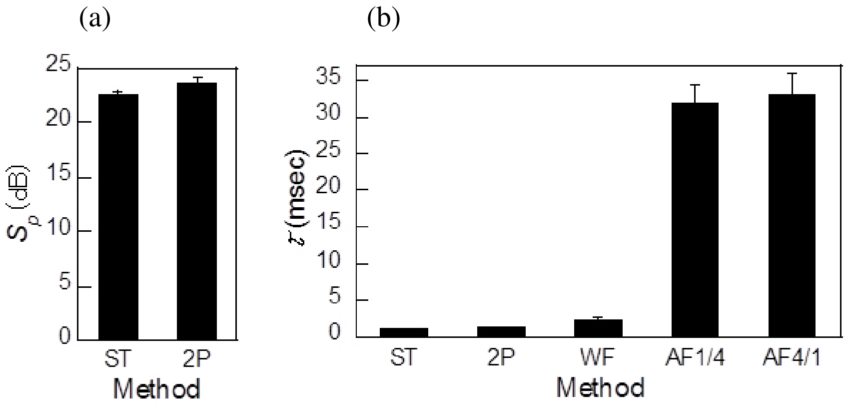


Fig. 1. (a) Mean peak sensitivities (absolute value) and (b) time constants τ of 25 listeners. ST: standard, 2P: two-point method of AM detection, WF: within-frequency, AF1/4: across-frequency gap detection with a 1000-Hz leading marker and a 4000-Hz trailing marker, AF4/1: across-frequency gap detection with a 4000-Hz leading marker and a 1000-Hz trailing marker. Error bars indicate standard errors of mean.

Table 1 presents the Pearson correlation coefficients among the various measures. Significant positive correlations were found between the two across-frequency gap thresholds and between τ obtained with the two AM methods. There were also significant positive correlations between τ and S_p obtained with the same method. S_p obtained with the standard method was significantly negatively correlated with the two across-frequency gap thresholds.

Discussion

The present results have several implications. First, the two-point AM method replicated the TMTF parameters obtained from the standard, multi-measurement AM method. The estimates of τ and S_p obtained with the two-point method were very close to and significantly positively correlated with those with the standard method. This extends the findings of Morimoto et al. (2018) who showed close resemblance of f_c obtained with these two methods. Together with its short measurement time, these results suggest that the two-point AM method is appropriate for clinical assessment of temporal resolution.

Second, the different measures of auditory temporal resolution, obtained with the AM detection and the within- and across-frequency gap detection, are dissociated from each other. While the time constants obtained with the two AM detection methods were similar to the within-frequency gap thresholds, they were not significantly correlated with each other. The across-frequency gap thresholds were much larger than and uncorrelated with any of the other measures of temporal resolution. Together with the close similarity and high correlation between the two across-frequency gap detection, the present results support the contention that different auditory processes underlie the within-frequency and across-frequency gap detection (Mori et al., 2015; Phillips et al., 1997).

Third, we found a significantly positive correlation between τ and S_p for each of the AM detection methods. For the two-point method, the significant correlation was mainly due to the method used to calculate τ, which included S_{p_2P} (Eq. (3)). This reasoning is supported by the insignificant correlation between τ and Δf_m, just noticeable modulation frequency, $r(23) = 0.05$. The significant correlation between τ and S_p in the standard method might have been an artifact caused by the fitting of a low-pass filter function to a limited number of points. This requires further investigation.

Finally, S_p was not significantly correlated with the PTA. S_p is often regarded as an index of intensity resolution (Shen, 2014; Shen & Richards, 2013). However, this might not have been true in the present study, in which the listeners' intensity resolution was assessed using the PTA. Unexpectedly, S_p obtained with the standard AM method was significantly negatively correlated with either of the two across-frequency gap thresholds. We have no ready answer for this and will look into it in a future study.

Table 1. Pearson correlation coefficients among various measures. S_{p_ST} and S_{p_2P} are shown in absolute values. PTA: pure-tone average of 500, 1000, 2000, and 4000 Hz.

	S_{p_ST}	S_{p_2P}	τ_{ST}	τ_{2P}	WF	AF1/4	AF4/1
S_{p_2P}	0.28						
τ_{ST}	0.49^{*}	0.53^{**}					
τ_{2P}	0.33	0.66^{***}	0.59^{***}				
WF	-0.22	0.03	0.02	0.08			
AF1/4	-0.55^{**}	-0.01	-0.26	-0.23	0.35		
AF4/1	-0.57^{***}	-0.03	-0.27	-0.25	0.30	0.84^{***}	
PTA	0.15	-0.09	-0.04	-0.07	0.05	-0.05	-0.15

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Acknowledgements

The experiment reported here was conducted by the second author as part of an undergraduate thesis submitted to Kyushu University. The research was supported by a Grant-in-Aid of the Japan Society for the Promotion of Science for Scientific Research (A) JP25240023 and Collaboration Research Fund of Rion Ltd. to S.M. Special thanks are due to Friedrich Müller for his support of the first author's participation in the 2018 Fechner Day meeting.

References

Eddins, D. A., & Green, D. M. (1995). Temporal integration and temporal resolution. In B. C. J. Moore (Ed.), *Hearing* (pp. 207–242). New York: Academic Press.

Formby, C., & Muir, K. (1988). Modulation and gap detection for broadband and filtered noise signals. *Journal of the Acoustical Society of America*, 82, 545-550.

Forrest, T. G., & Green, D. M. (1987). Detection of partially filled gaps in noise and the temporal modulation transfer function. *Journal of the Acoustical Society of America*, 82, 1933-1943.

Ito, K., Takamura, A., & Mori, S. (2016). Auditory across-channel processing as assessed by monaural and between-ear temporal gap detection tasks. *Journal of the Acoustical Society of America*, 140, 3328.

Mori, S., Oyama, K., Kikuchi, Y., Mitsudo, T., & Hirose, N. (2015). Between-frequency and between-ear gap detection and their relation to perception of stop consonants. *Ear & Hearing*, 36, 464-470.

Morimoto T., Irino, T., Harada, K., Nakaichi, T., Okamoto, Y., Kanno, A., Kanzaki, & S., Ogawa, K. (2018). Two-point method for measuring the temporal modulation transfer function. *Ear & Hearing*.

Oxenham, A. J. (2016). Predicting the perceptual consequences of hidden hearing loss. *Trends in Hearing*, 20, 1-6.

Phillips, D. P., & Smith, J. C. (2004). Correlations among within-channel and between-channel auditory gap detection thresholds in normal listeners. *Perception*, 33, 371–378.

Phillips, D. P., Taylor, T. L., Hall, S. E., Carr, M. M., & Mossop, J. E. (1997). Detection of silent intervals between noises activating different perceptual channels: Some properties of “central” auditory gap detection. *Journal of the Acoustical Society of America*, 101, 3694–3705.

Plack, C. J., Barker D., & Prendergast, G. (2014). Perceptual consequences of “hidden” hearing loss. *Trends in Hearing*, 18, 1-11.

Shen, Y. (2014). Gap detection and temporal modulation transfer function as behavioral estimates of auditory temporal acuity using band-limited stimuli in young and older adults. *Journal of Speech Language and Hearing Research*, 57, 2280-2292.

Shen, Y., & Richards, V. M. (2013). Temporal modulation transfer function for efficient assessment of auditory temporal resolution. *Journal of the Acoustical Society of America*, 133, 1031-1042.

JUDGING NOISE ANNOYANCE AS A FUNCTION OF TASK LOAD: EFFECTS OF NOISE SENSITIVITY

Wolfgang Ellermeier[1], Florian Kattner[1], Ewald Klippenstein[1], and Catherine Marquis-Favre[2]
[1] Institut für Psychologie, TU Darmstadt, Alexanderstr. 10, 64283 Darmstadt, Germany
[2] Université de Lyon, Ecole Nationale des Travaux Publics de l'Etat (ENTPE), Laboratoire Génie Civil et Bâtiment, 3 Rue Maurice Audin, F-69518 Vaulx-en-Velin, France
ellermeier@psychologie.tu-darmstadt.de, kattner@psychologie.tu-darmstadt.de, catherine.marquisfavre@entpe.fr

Abstract

In order to study noise annoyance as a nuisance to some focal activity, listeners were exposed to short 5-s recordings of road-traffic noise while performing a visual multiple-object tracking task of approximately the same duration. The noise samples were repeatedly presented at levels ranging from 50 to 70 dB(A) as the difficulty of the primary task varied from tracking between 2 and 4 of 16 objects simultaneously presented on the screen. At the end of each trial, subjects had to make a response to the object-tracking task, immediately followed by an annoyance rating for the sound on a visual-analogue scale. Twenty-nine subjects participated, being exposed to all combinations of primary task difficulty and SPL in random order, while their skin conductance response was measured. Psychophysical annoyance-scaling functions were related to the participants' noise sensitivity: On average, more sensitive individuals tended to assign slightly higher ratings to the more intense stimuli. The magnitude of the skin conductance response increased systematically with the sound pressure level of the annoying road-traffic noise presented on a given trial. That way, interrelations between noise sensitivity, noise-annoyance, task load during exposure, and concurrent psychophysiological responses were analyzed.

Noise annoyance is typically studied 'in the field', i.e. in surveys querying a large number of participants on the inconveniences they experienced due to particular noise sources (e.g. aircraft noise) over a relatively long period of time, such as a year. This kind of survey methodology has been standardized to some extent (Fields et al., 2001).

By contrast, the validity of studying noise annoyance in the laboratory has been questioned on principled grounds (e.g. Guski and Bosshardt, 1992). The concern is that people – removed from their habitat – might tend to simply judge loudness, aesthetic preference, or some other salient perceptual feature of the sounds while disregarding the disruptive nature and affective component implied in the notion of annoyance. Nevertheless, it is generally agreed upon that short-term effects of noise may be studied under laboratory conditions by having participants imagine a situation like relaxing at home and then judging noise samples played back as authentically as possible (e.g. Gille et al., 2016; Ryu and Jeon, 2011).

In the present study we chose to have participants actually work on a primary task, not requiring a situation to be ‘imagined’, while being exposed to irrelevant and potentially annoying traffic noise. The goal of the study was threefold: (1) to determine whether noise annoyance may be reliably judged as a ‘secondary task’ while the subject is occupied with some primary task of varying difficulty, (2) to assess whether skin conductance responses (SCRs) to isolated noise events may serve as a valid indicator of annoyance and (3) to assess the role of noise sensitivity in modulating both annoyance ratings and psychophysiological responses.

Method

Participants

A sample of N = 29 participants (18 men, age range 18-61 years, MD = 22) was recruited from the TU Darmstadt student body and members of the laboratory. All participants claimed to have normal hearing, except for two who reported occasional tinnitus not affecting their hearing sensitivity. The undergraduate students participated for course credit.

Stimuli

Three recordings of vehicle pass-by noises prepared for an earlier study (Marquis-Favre and Morel, 2015) were used as stimuli. They had been recorded using stereo microphones in an ORTF arrangement and were between 3.7 and 6.4 s long. The recordings were of a motorcycle decelerating and breaking while passing by, and of two trucks, one accelerating, the other one decelerating and eventually idling near the recording site. Thus the recordings were sufficiently rich in timbre and spectral change over time to be distinguishable and potentially elicit different degrees of annoyance. The levels recorded at a distance of 3 m when passing the microphones were approx. 70 dB(A) for the accelerating truck sound and 62 dB(A) for the remaining two sounds. In the present experiment, in order to vary sound pressure level, they were presented at these (original) levels and 6 lower levels, in 2-dB(A) steps, thus resulting in ranges 50-62 (motorcycle, decelerating truck) and 58-70 dB(A) (accelerating truck).

Apparatus

The stereo recordings were presented via electrodynamic headphones (Beyerdynamic DT 990) without further processing. To that effect they were D/A converted (with 44.1 kHz, 32 bits) by a high-quality sound card (RME multiface II) and passed through a headphone amplifier (Behringer Pro 8). The playback at different levels was achieved by storing attenuated copies of the sounds. Levels were verified 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).

During the listening test, skin conductance was continuously measured by means of a psychophysiological recording system (Biopac MP 150WS-NDT and GSR100C module with AcqKnowledge 3.9 data acquisition software) via two Ag/AgCl electrodes (diameter 8 mm) filled with isotonic gel (0.5% saline in a neutral base) attached to the palm of the participant's non-dominant hand. Skin conductance was recorded at a sampling rate of 250 Hz.

Procedure

On each trial, participants had to perform a visual-attention task while being exposed to one of the traffic-noise recordings. The task chosen was a 'multiple object tracking' (MOT) task (Pylyshyn and Storm, 1988). At the outset of each trial, participants were presented with an array of 16 circles that were moving in random directions within a grey circular field occupying almost the entire screen (details as in Green and Bavelier, 2006). A subset of 2, 3, or 4 circles was presented in orange color (i.e., the target circles), while the remaining circles were presented in yellow (i.e., the distractor circles). Participants were instructed to track the orange target circles throughout the trial. The target circles remained orange for 2 seconds, after which they changed color to yellow for 4 seconds. The participants were instructed to continue tracking the circles that initially were orange, thus the total tracking time per trial was 6 seconds. At the end of each trial, a single circle in the display changed color to white (i.e., the probe circle). Participants were instructed to indicate whether this probe circle was one of the (originally orange) target circles by pressing either the left mouse button for "yes", or the right mouse button for "no". The probe circle was originally a target circle on 50% of the trials, and a distractor circle on 50% of the trials. Participants received feedback on their performance (correct / false) immediately after they had made a response.

After a 1-s delay, a horizontal visual-analogue scale appeared on the screen prompting the participant to indicate how annoying he or she found the sound played during that trial. The scale was labelled "Gar nicht lästig" (not annoying at all) on its left and "Extrem lästig" (extremely annoying) on its right end.

Each type of trial (3 sounds x 7 levels x 3 task difficulties x 2 types of probes) was presented only once to each participant, in random sequence, thus resulting in a total of 126 trials. The entire procedure, including preparing the skin conductance recordings and filling out a noise sensitivity questionnaire (LEF, Zimmer & Ellermeier, 1998) took approximately 1h.

Results

Annoyance ratings

The VAS annoyance ratings (expressed as a proportion of the scale 0-1.0) were arithmetically averaged across subjects and appropriate conditions. Performance in the MOT task was evaluated as the proportion of accurate responses. Fig. 1 (left graph) shows that accuracy drops as the task difficulty is increased from tracking 2 to 4 circles. That is evidence that subjects attended to the task, performed well above chance level, and were affected by the cognitive load it imposed. Noise sensitivity did not have a significant effect on performance in the visual-attention (MOT) task.

Fig. 1 (right graph) shows mean annoyance ratings as a function of A-weighted sound pressure level, contrasting participants of low self-reported noise sensitivity (n_{low} = 15, open symbols) with those reporting high noise sensitivity (n_{high} = 14, filled symbols), based on a median split. It is evident that the more noise-sensitive individuals tend to produce higher annoyance ratings throughout the range of levels studied. This effect is not statistically significant, however.

The difficulty of the primary visual-attention (MOT) task, however, did have a small, but significant, effect (not shown in Fig. 1) on annoyance ratings: The more demanding that task was on a given trial, the higher the reported annoyance. With two circles to track, mean annoyance was M=0.36, SD=0.18; with three circles to track it was M=0.38, SD=0.18, and at the most difficult level, with four circles, M=0.41, SD=0.19. This was reflected in a highly significant main effect of difficulty in a 3 (difficulty) x 11 (sound level) x 2 (noise sensitivity groups) mixed-factors ANOVA, $F(2,54)=13.08$; $p<.001$; $\eta^2=.009$ in which difficulty and sound level constituted within-subjects factors. The main effect of sound level on annoyance ratings was also highly significant, $F(10,270)=46.06$; $p<.001$; $\eta^2=.19$, while the main effect of noise sensitivity was not, $F(1,27)=.39$; $p=.53$. None of the interaction effects turned out to be significant.

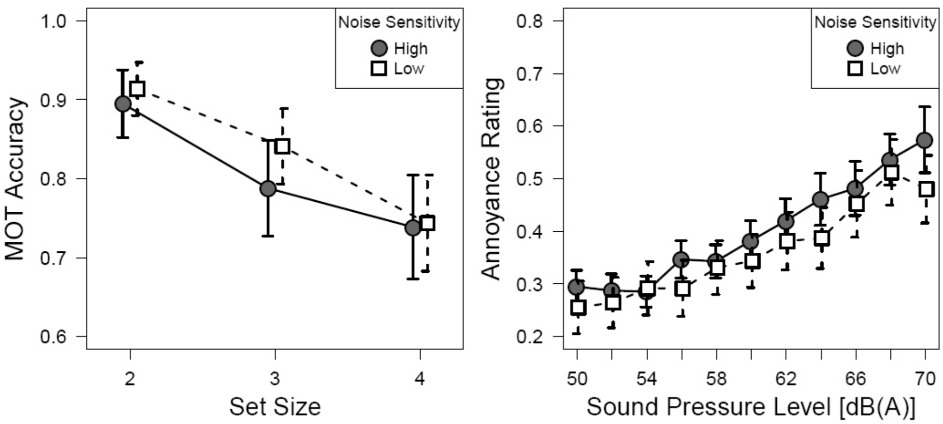


Fig. 1. Left: Mean accuracy in the visual attention task as a function of set size (the number of objects to be tracked). Right: Mean annoyance ratings as a function of SPL. Both graphs show separate curves for the group of highly noise sensitive participants (n = 15) contrasted with those of low noise sensitivity (n = 14).

Electrodermal responses

To analyze the skin conductance recordings, they were first temporally aligned with the 6-s trial structure of the visual attention task performed under noise. To that effect, the change in skin conductance, expressed in μS and referenced to the conductivity level at the onset of the noise (and of the MOT task) was averaged across trials accompanied by the same sound pressure level and tracked for the 6 s of the MOT task and the subsequent response interval.

The result is seen in Fig. 2 which shows the magnitude and duration of the skin conductance response (SCR) to increase with sound pressure level; the effect is most marked for the four highest SPLs which are due to the accelerating truck sound played back at levels of up to 70 dB(A). This effect is statistically significant when the ‘first-interval response’ (FIR, i.e. the average skin conductance in seconds 1-4 of the SCR which is assumed to reflect an orienting response; see Koenig et al., 2017) is analyzed, $F(10,260)=3.24$; $p<.001$; $\eta^2=.03$.

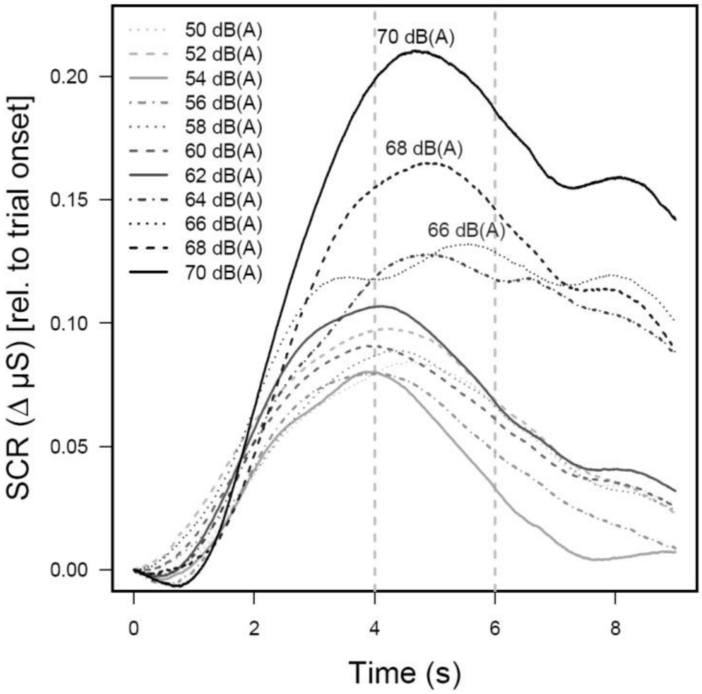


Fig. 2. Mean change in skin conductance as a function of sound level. Curves are based on 18 trials per level and participant and averaged across 28 subjects for whom valid EDA data were obtained.

Discussion

The results of the present pilot study show that annoyance judged while performing an absorbing task is significantly – though to a small extent - influenced by the cognitive load imposed by that task. Furthermore, measuring event-related electrodermal responses appears to be a valid measure of perceived annoyance, exhibiting an unequivocal level dependence. Finally, noise sensitivity appears to play a role in augmenting both annoyance ratings and the magnitude of psychophysiological responses, but that will have to be confirmed in a larger sample. All of these findings are of limited value before an actual comparison with a more conventional assessment of annoyance in the absence of a distracting task is performed.

References

Fields, J.M., De Jong, R.G., Gjestland, T., Flindell, I.H., Job, R.F.S., et al. (2001). Standardized General-Purpose Reaction Noise Questions for Community Noise Surveys: Research and a Recommendation. *Journal of Sound and Vibration*, *242*, 641-679.

Gille, L.-A., Marquis-Favre, C., & Klein, A. (2016). Noise annoyance due to urban road traffic with powered-two-wheelers: quiet periods, order and number of vehicles. *Acta Acustica united with Acustica*, *102*, 474-487.
Green, C.S., & Bavelier, D. (2006). Enumeration versus multiple object tracking: the case of action video game players. *Cognition*, *101*(1), 217-245.
Guski, R., & Bosshardt, H.-G. (1992). Gibt es eine „unbeeinflusste" Lästigkeit? [Is there an "unbiased annoyance"?] *Zeitschrift für Lärmbekämpfung*, *39*(3), 67-74.
Koenig, S., Nauroth, P., Lucke, S., Lachnit, H., Gollwitzer, M., & Uengoer, M. (2017). Fear acquisition and liking of out-group and in-group members: Learning bias or attention? *Biological Psychology*, *129*, 195-206.
Marquis-Favre, C., & Morel, J. (2015). A simulated environment experiment on annoyance due to combined road traffic and industrial noises. *International Journal of Environmental Research and Public Health*, *12*, 8413-8433.
Pylyshyn, Z., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*, 179-197.
Ryu, J.K., & Jeon, J.Y. (2011). Influence of noise sensitivity on annoyance of indoor and outdoor noises in residential buildings. *Applied Acoustics*, 72, 336–340.
Zimmer, K., & Ellermeier, W. (1998). Konstruktion und Evaluation eines Fragebogens zur Erfassung der individuellen Lärmempfindlichkeit [Construction and evaluation of a questionnaire for measuring individual noise sensitivity]. *Diagnostica*, 44, 11-20.

BINAURAL LOUDNESS OF MOVING SOURCES IN FREE FIELD: PERCEPTUAL MEASUREMENTS VERSUS AT-EAR LEVELS

Sabine Meunier, Sophie Savel, Jacques Chatron, and Guy Rabau
Aix Marseille Univ, CNRS, Centrale Marseille, LMA, Marseille, France 4 impasse Nikola Tesla, CS 40006, 13453 Marseille Cedex 13

Abstract

Most investigations on the variations of loudness with the spatial position of a sound source have been made for static sounds. The purpose of this work was to study the loudness of a moving source. By analogy with studies on difference in loudness between sounds increasing or decreasing in intensity (without movement of the source), we studied the global loudness of a moving sound. The analogy with the sounds whose intensity varies is direct because the at-ear level depends on the position of the source, so a moving sound will create levels that vary over time at the entrance of the auditory canal. We measured the overall loudness of a moving source as a function of the starting and ending positions of the stimulus and of its direction of rotation. Overall, we did not find any overall loudness difference according to the direction of variation of the source. Moreover, the results obtained with a static sound seem to confirm, with absolute magnitude estimation, the amount of directional loudness sensitivity measured previously with an adaptive method.

In free field, loudness depends on the position of the sound source (Sivonen and Ellermeier, 2006). In order to quantify the effect of the incidence angle on loudness, the directional loudness sensitivity (DLS) is measured. DSL is defined as the level difference required for equal loudness between a frontal reference sound (azimuth 0°, elevation 0°) and a test sound at a given position. A negative DLS means that the test sound has been perceived softer than the frontal sound and vice versa. In a previous study, we showed a decrease in DLS with an increase in azimuth of an amount of more than 10 dB on average (25 listeners, Meunier et al., 2016). Different studies have also examined the loudness of sounds that increase and sounds that decrease in level. For sounds that only differ in temporal envelop, it has been shown that the global loudness of a sound whose level increases is greater than the global loudness of a sound whose level decreases (Ponsot et al., 2015a, 2015b). This phenomenon has been called asymmetry in loudness.

When a sound source is moving around a listener, the at-ear level of the sound varies. Moreover, if we refer to the studies on directional loudness, its loudness should also vary. The aim of the work presented here was to explore how global loudness of moving sounds is formed and the main point was to determine whether there is an asymmetry between sounds that move in opposite directions as their level and loudness also vary in opposite directions.

Experiment

Ten normal-hearing listeners participated in the experiment.

The loudness of moving sounds was measured in an anechoic room. The Vector Amplitude Base Panning (VBAP) method was used (Pulkki, 1997). It consisted in creating virtual sources from several loudspeakers situated at equidistance from the listener head in varying the gain of each loudspeaker. With this technique it was possible to move the sound around the listener. The speed of the movement was constant. The stimuli used for the experiment were a third-octave noise band centered at 5 kHz and white noise (140-17 000 Hz) of duration 2 s. The levels were 45, 50, 55, and 60 dB SPL. They were measured in the absence of the listener at the theoretical center of the head. The experiment was realized in the horizontal plan (elevation = 0°). The azimuths of the third-octave-band stimulus were either fixed at 75 or 135° (static sources) or varied from 75 to 135° and inversely (moving sources). For the white noise, the static positions were 45, 75, 135 and 150°, and the sound moved from 45 to 150° (and inversely) and from 75 to 135° (and inversely). These positions were chosen in order to produce different at-ear levels and different loudness levels when the stimulus varied in azimuth. For the third-octave-band stimulus, the directional loudness sensitivity decreased by about 8 dB from 75 to 135°; the at-ear level decreased by about 9.5 dB on the right ear (ipsilateral) and varied by less than 2 dB on the left ear (contralateral). For the white noise, the directional loudness sensitivity decreased by about 2.5 dB from 45 to 150° and by about 2 dB from 75 to 135°. The at-ear level decreased by about 6 dB on the right ear from 45 to 150° and by about 4 dB from 75 to 135°. It varied by about 4 dB on the left ear from 45 to 150° and by less than 1 dB from 75 to 135°. The at-ear levels were measured with blocked ear canals for all loudspeaker positions and all listeners. The values given above correspond to the average of the measurements made previously (Meunier et al., 2016) on the 10 listeners of the present work. The loudness was evaluated using an absolute magnitude estimation procedure. One block consisted of all conditions for one stimulus (static and moving sources at the 4 different levels). Each condition was repeated 3 times. A block was repeated once. Thus, for one listener and one condition the loudness was the geometric mean of 6 estimates. A training block was also run for each stimulus.

Results

The geometric means of the 10 listeners are shown in figures 1 and 2. There was no significant effect of the direction of variation on the estimates either for the third-octave-band stimulus [$F(9,1)=0.013$, $p=0.91$] or for the white noise [$F(9,1)=3.33$, $p=0.1$ for 45-150° ; $F(9,1)=0.32$, $p=0.58$ for 75-135°].

For the third-octave-band, a post-hoc LSD showed significant differences between the estimates of the static sounds at 75° and the other conditions ($p<0.001$) and no significant differences between the three other conditions. Whatever the direction of movement, the estimates of the moving sounds corresponded to the estimate of the sound at the position of 135° which was the softest sound.

For the white noise, there were no significant differences between all conditions.

For the static sounds, we found almost the same difference in level between sounds of different positions that were equally loud using an absolute magnitude estimation method (figures 1 and 2, left panels: 6 dB for the third-octave-band, 2 dB for the white noise between 45 and 150° and 1.8 dB for the white noise between 75 and 135°) and using an adaptive method to measure the DLS (figures 1 and 2, right panels : 7 dB for the third-octave-band, 2.5 dB for the white noise between 45 and 150° and 2 dB for the white noise between 75 and 135°).

The exponents of the loudness functions seem to be independent of the sound source position and of the movement (figures 1 and 2, left panels). They were around 0.25 for the third-octave band stimulus and around 0.29 for the white noise.

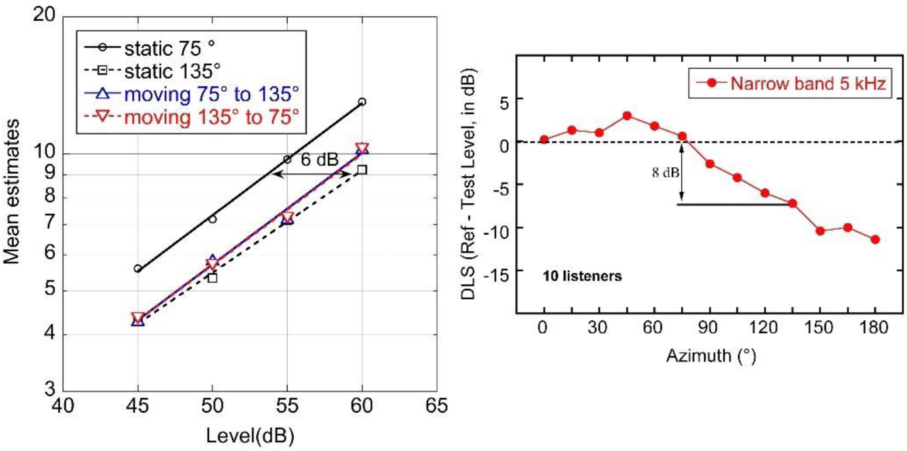


Figure 1. Geometric means (10 listeners) of the loudness estimates for the third-octave band centered on 5 kHz for static and dynamic sounds (left panel) and Directional Loudness Sensitivity as a function of the azimuth: mean of 10 listeners for the third-octave band centered on 5 kHz (right panel)

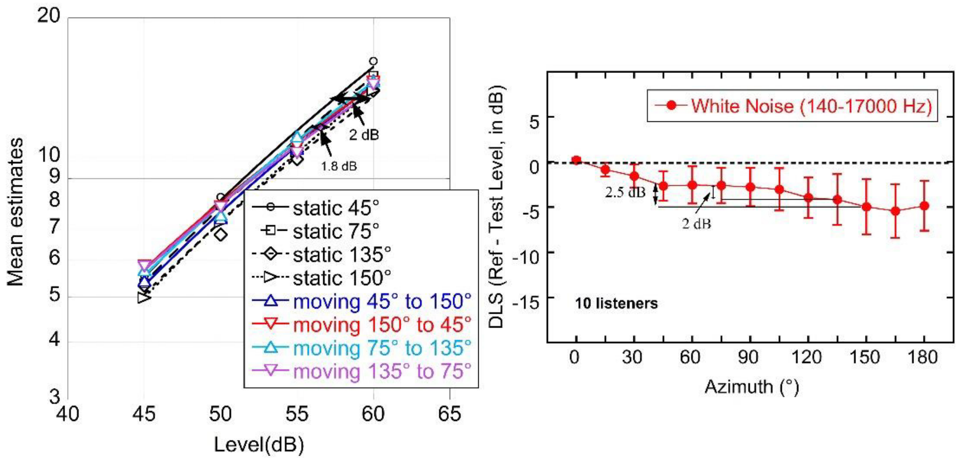


Figure 2. Geometric means (10 listeners) of the loudness estimates for the white noise for static and dynamic sounds (left panel) and Directional Loudness sensitivity as a function of the azimuth: mean of 10 listeners for the white noise (right panel)

Discussion and Conclusion

The direction of the movement does not seem to affect the global loudness while DLS and at-ear levels vary in opposite ways for opposite directions. We did not find any asymmetry in loudness in this case, even if the at-ear level varied. This finding is in disagreement with the results of studies on asymmetry in loudness, in which the global loudness of an increasing sound was found larger than that of a decreasing sound. This may be due to the small amount of at-ear level variation (9.5 dB for the third-octave-band stimulus, 6 dB for the white noise from 45 to 150° and 4 dB from 75 to 135°, in the ipsilateral ear), much smaller than the dynamics of the stimuli used in experiment on loudness asymmetry for static sounds (15 or 30 dB). It may also be due to an influence of the contralateral ear in which the level variations were very small but were present.

For the third-octave band, we found than the loudness of the moving sound corresponded to the loudness of the sound at 135°, whatever the direction of variation. This result is in discrepancy with a prevalence of the louder part of the sound on the global loudness because at 135° the loudness of the third-octave-band stimulus is softer than at 75°. This phenomenon might be due to the fact that at 135° the source is not in the visual field of the listener, which would induce more attention to that part of the space. But for narrow-band noise, the localization is very poor, and around 5 kHz the sound would be localized in front of the listener (Blauert, 1983).

Based on a different procedure, this experiment confirms the amount of directional loudness sensitivity found previously (Meunier et al., 2016). The absolute magnitude estimation is much faster than the adaptive procedure used in Meunier et al. (2016), and would be a promising alternative for measuring directional loudness sensitivity.

The exponents of the loudness functions were the same for the static and dynamic stimuli and did not depend on the position of the sound source. However, the exponents were very small, much smaller than the ones found usually in literature. For example, Canévet et al. (2003) found an exponent of 0.49 for a 4 kHz pure-tone. At 1 kHz the exponent is about 0.6 (Steven, 1957). This may be due to a contextual effect as static and dynamic sounds were presented in the same blocks.

Acknowledgements

This work was supported by the LoudNat project funded by the French National Research Agency ANR-11-BS09-016-01.

References

Blauert J. (1983). Spatial hearing.The psychophysics of human sound localization. MIT Press.

Canévet G., Teghtsoonian M., and Teghtsoonian R. (2003). Assimilation and asymmetry of loudness change in magnitude-estimation measurements. Acta Acustica united with Acustica, 89, 530-539.

Meunier, S., Savel, S., Chatron, J., and Rabau, G. (2016). Interindividual differences in directional loudness. J. Acoust. Soc. Am., 140, 3268–3268.

Ponsot, E., Susini, P., and Meunier, S. (2015). A robust asymmetry in loudness between rising- and falling-intensity tones. Atten. Percept. Psychophys., **77**, 907–920.

Ponsot, E., Susini, P., and Meunier, S. (2015). Erratum to: A robust asymmetry in loudness between rising- and falling-intensity tones. Atten. Percept. Psychophys., **77**, 921–921.

Pulkki, V (1997). Virtual sound source positioning using vector base amplitude panning. J. Audio Eng. Soc., 45(6) :456–466.

Sivonen, V. P., and Ellermeier, W. (2006). Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation. J. Acoust. Soc. Am., 119, 2965–2980.

Steven S. S. (1957). Concerning the form of the loudness function. J. Acoust. Soc. Am., 29, 603-606.

EFFECT OF SOUND ON MEMORIZATION OF VISUAL IMAGES

Natalia Postnova[1], Shin-ichiro Iwamiya[2], and Gerard B. Remijn[3]
[1] *Human Science Course, Graduate School of Design, Kyushu University, Fukuoka, Japan*
[2] *Department of Communication Design Science, Kyushu University, Fukuoka, Japan*
[3] *Department of Human Science / Research Center for Applied Perceptual Science, Kyushu University, Fukuoka, Japan*
nataliapostnova@gmail.com, remijn@design.kyushu-u.ac.jp

A series of experiments on the effect of sound on memorization of visual images was conducted to examine the effect of sound on the test performance in general and on memorization of individual visual images. Our recent experiment replicated and confirmed the previous findings (Postnova & Iwamiya, 2017) under more controlled conditions. The experiment consisted of a memory-span test, in which sequences of black-and-white visual images were presented under three different conditions. In two conditions the visual images were accompanied by a sound, a 440 Hz short pure tone. The third condition was a no-sound control condition. In one of the two sound conditions, a sound was played with every image in the sequence synchronized with the appearance of the image. In the other sound condition, only half of the images in the sequence were accompanied by the sound. The effect of sound on image memorization was examined by comparison between performances in these three conditions. The one-way repeated measures ANOVA did not show a significant difference in image recall. However, in the condition in which half of the images was accompanied by the sound, we found that among correctly recalled images, the number of recalled images was significantly higher ($p<0.01$) with sound than without sound. We will also present data in which we try to confirm and clarify these findings by conducting an experiment with less meaningful images, using a wider variety of sound conditions.

Reference

Postnova, N. & Iwamiya, S. (2017). Effect of sound on memory and impressiveness of visual imagery. *Fechner Day 2017* (pp. 77-83). Fukuoka, Japan.

SESSION II

FROM THE OUTER PSYCHOPHYSICS OF SS STEVENS TO THE INNER PSYCHOPHYSICS OF MIND GENOMICS

Howard R. Moskowitz
Mind Genomics Associates, Inc., White Plains, New York, USA
mjihrm@sprynet.com

Abstract

This paper combines the author's reminiscences of a life in psychophysics with the creation and development of Mind Genomics, a new science which ideas, and which reveals new-to-the-world groups of people with different viewpoints about the same idea. The paper shows the method, presenting a range of applications from products to the law to social issues. The paper finishes with a view of the prospects of Mind Genomics as a science, and the possible next stage in the evolution of psychophysics from an Outer Psychophysics to an Inner Psychophysics.

A brief introduction

It is the summer of 2018. Just about 50 years ago, the writer sat in his office in room 910 of William James Hall, Harvard University, epicenter of the modern-day psychophysics, preparing to show his 'data' to SS (Smitty) Stevens, the eminence grise of psychophysics, and Howard's not-really-acknowledged advisor. Why Smitty was, Howard's advisor, de facto, but would not acknowledge that role, turns out to be more complex, pointing to an interplay of personalities, mutual respect, fear, and the utter unknowability of this grand old man. This 'denial' of academic paternity in a formal sense, but acceptance in every-day behavior and during the final oral exams, is also important for this development of the Inner Psychophysics. It would spur the author to move on to both gain acceptance from Stevens as a professional, yet 'break away' to establish himself as a separate being, in a classic pattern of child development.

This paper is a short memoir, a history, written from the perspective of the author, who grapples daily with the seeming minor issue of '*how did I get here*?' That's not the issue at all, not really. Rather, the issue is what do all these experiments really mean, not so much for science, as for this particular individual, Howard Moskowitz, confronting his chosen, possibly even beloved discipline, psychophysics.

Harvard's psychophysics, circa 1967

To understand this paper, it is worth understanding the intellectual atmosphere of Harvard Psychophysics, circa 1967. Smitty was, of course, THE PROFESSOR. Gruff, dour, given to an occasional smile when he sat reading the Wall Street Journal in the early morning, talking to a non-student, e.g., a young wife of a student. To the acolytes of psychophysics, those two or three of us by that time, Smitty was less than charming and delightful, more like a blowtorch, singing away anything that was not intellectually fireproof.

There was a deep, underlying current, of a material universe, philosophically akin to some of the French philosopher La Mettrie's sense of a god who created the world, and left the world to let it move along in clockwork precision. There were rules to be learned about this magnificent clock. And in Smitty's mind, the notion was 'nomos,' rule, invariance, the surprisingly simplicity of the world. In other words, to Smitty, there was a mind, transforming physical stimuli to subjective percepts, according to rules, rules which were invariant. To Smitty, therefore, and thus to us graduate students, the goal of science was to uncover this simple set of rules, some of which would manifest themselves in lawful, reproducible, invariant relations between measurable physical stimuli such as the sound pressure of tones, the molarity of sugar solutions, and the consequent perception, the loudness of the tones the sweetness of the sugars. All in all, a mechanical universe. The role of science was simply to find a rule. Not 'the rule,' of course, but 'a first order approximation,' to use one of Smitty's favorite phrases. Parenthetically, that was not only a phrase, but a marching shout to accompany one into a battle with nature to pry open her secrets.

The formative encounter with Smitty Stevens

Mind Genomics, my suggestion that it is Inner Psychophysics, did not grow out of a calm, intellectual path, of the type that one hears about in today's graduate programs, what with nurturing professors, study groups, and the like. The encounters with Stevens came about because the author was and remains singularly inept when it comes to equipment. In the psychophysics of the 1960's, a great deal of attention was paid to the equipment which that would produce the stimulus. Indeed, some of the joy of the field arose out of putting together equipment which worked. Indeed, equipment was as much the ground substance of psychophysics as was the experimenter. Smitty's words of June, 1967, 51 years ago, remain in my ears… '*Howie...you are not an elektroniker...why don't you do something that you can master, such as taste or political scaling*.'

The die was cast. I would work in the sense of taste, where I did not have to plug in equipment, read user manuals, and spend all my time on externalia, the equipment. Rather, I would be more comfortable because I could spend my time and emotional energy on internalia, the thinking, the pondering about the study. It may well be that moving away from equipment to simple paper-portion control cups by Lily, the cup manufacturer, would eventually push me into this world of Inner psychophysics. I didn't have to worry about calibrating my equipment, nor about making fine measurements. No one knew much about taste. All I had to do was a simple experiment, creating mixtures of tastants of different qualities (e.g., sugar and salt), vary the components in systematic concentrations, and obtain ratings of taste intensity. I was free to ponder. What would happen to the relation between rated sweetness and sugar concentration, when I threw in a constant amount of salt, sodium chloride? Were the relations something which could win me a PhD? Simple experiments these, but in retrospect so powerfully different as to change the course my life.

Taste mixtures

One of the 'nice things' about taste work is that I had to create the test stimuli by hand. No electronics, no sophisticated equipment, no psychological distance from the stimuli, no emotional distance from the stimuli. Rather, an immediacy with one's stimuli, and an immediacy of collecting the data and plotting it. Oh, and one more thing. The respondent had to focus on sensory intensity. Some of the mixtures tasted good, some tasted downright terrible, and still others tasted 'strange,' especially those mixtures comprising low concentrations of both sugar and salt, respectively.

But what does it taste like, really? It tastes terrible.

Smitty's 'Outer Psychophysics' focused on the rather soul-less, almost wraithlike pattern of behavior. Psychophysics in those days reveled when one plotted the data to uncover straight lines in log-log coordinates with the physical measure of intensity on the abscissa and the median rating of perceived intensity on the ordinate. It was heady to do these experiments, to see these straight lines with reproducible slopes emerging when one plotted the data, which happened every night. The PhD slowly progressed, one taste at a time, an expectorated (i.e., spit) taste solution moving from the mouth of the observer to the crusty large pot which contained the expectorated solutions, later to be emptied, washed, and used for the next respondent or O (observer in the language of the psychophysicist.) I was doing research, doing Outer Psychophysics, plotting relations between variables, seeing nature's secrets unfold in front of my youthful, excited, occasionally overawed eyes.

Or was I? It's one thing to measure the loudness of sounds, the brightness of lights, even the pain of electric shock, not to mention the hardness of rubber, the area of circles, and occasionally the odor intensity of some chemical. The stimuli were distal, created by machine, sent to the observer, processed, disappearing as quickly as they arrived, with numbers, magnitude estimates, left in their place. This was Outer Psychophysics, 'outer' not only in a philosophical sense, but outer because there was no intimate contact of the person with the stimulus.

Taste was altogether different. Taste was intimate, a liquid rolling around in the mouth, a liquid which more often than not was liked and disliked, reactions to this stimulus coming faster than the rather intellectualized response, 'sensory intensity.' It was that intimacy with the stimulus, that insistence of hedonics, liking and disliking, coming to the surface to announce itself, which pushed me towards what would become Inner Psychophysics. But not yet. The path was just beginning, least the interesting part to me, anyway.

Looking back to this formative period, it is becoming increasingly clear to me that hedonics, likes and dislikes, so inherent in taste and smell, was slowly moving me from being a classical psychophysicist to one interested in the mind, as well as interested in the disciplined search for orderly relations between variables. It is one thing to understand the dynamics of hearing and seeing, where there is no 'hedonics' raising its head, no dimension of liking/disliking. With seeing and hearing, one focuses on orderly relations between variables. People should all judge the perceived loudness in the same way. The variability across people is random error. Psychophysics is thus just another science, looking for invariances, unchanging rules. Liking, hedonics, throws a 'monkey wrench' into this perfectly operating mechanical world. People can react accurately to the stimuli, rating say the perceived sweetness of sugar solutions of different molarities so that the ratings 'track' the physical levels. Look away for a moment, and in rushes hedonics to muddy up the waters. People whose judgments of intensity would give delight to the psychophysics because they all 'tracked the concentration in the same way, by their ratings,' could then turn around and disappoint. For high levels of sucrose in solution, very sweet, almost syrupy solutions, some individuals would like this very sweet solution, others would dislike it (Ekman & Akesson, 1965.)

If hedonics of responses to sweet solutions managed to disturb the nice invariance sought by psychophysics, a more disturbing results was reported by author Moskowitz, et. al., (1975), working with Indian laborers of Karnataka origin, with the experiment done a few years before at St. John's Medical College. These laborers were able to judge the intensities of the basic tastes, sweetness, saltiness, bitterness, sourness just as accurately as did occidental respondents. So far, so good for psychophysics. But enter judgments of liking. These laborers loved increasing concentrations of acid, unlike the response of everyone else, who disliked increasing concentration of acid.

And so, mind enters, disturbing the beautiful invariance of Outer Psychophysics. The respondents all judged the sensory intensities in similar ways, correcting of course for scale usage, but, but differed in how much they LIKED what they tasted.

The initial forays into liking, discouraged at Harvard by Smitty, was welcomed at the US Army Natick Laboratories, by Harry Jacobs, Linda Bartoshuk, Herbert Meiselman, colleagues of Howard during the early days. What was important was that psychophysics now had something to do with the mind. During those formative, early years, there was no excitement at the thought that this research was dealing with the Inner Psychophysics, as the author now suggests in this paper. Hedonics, likes and dislikes were simply 'here.' They were topics to be studied by young, 20-something researchers, at the start of their careers.

The forays into hedonics, liking and disliking, produced the necessary fertilizer to grow the seedlings of Mind Genomics. The original work was not mine, but rather belongs to researchers a century ago, Engel (1928), writing in German a decade after World War I, studying votes by panellists of 'like or dislike' versus the concentration of a tastant, e.g., sugar or salt of various concentrations in water. Engel does not seem to have had an intellectual connection to the world of psychophysical scaling, but rather used the more traditional method of measuring the percent of times a test stimulus was deemed pleasant or unpleasant, and how the vote for 'pleasantness' changed with changing concentration. So here was taste psychophysics in its early days already grappling with this internal, intractable phenomenon known as liking, a phenomenon which insistently pointed to an Inner Psychophysics, for by what other process would we have people perceiving the sensory intensity in the same way as a first approximation, anyway, as far as we can measure it. And yet these same people responding so differently when the criterion was liking?

From the tongue to the mind & Mind Genomics - Colgate Dental Cream & Court Shepard

My 'epiphany' occurred in early 1980, at a meeting in Toronto, Canada. By 1980, I had been comfortably ensconced in the market research business, testing products, optimizing products, and in general trying to produce the best possible product for the client. By clients I mean companies as diverse as Pepsi Cola, Unilever, Armour-Dial, and so forth. My business consisted almost entirely of the testing of, and consulting about, food and other fast-moving consumer goods. And, of great importance, was the reality that I was in the 'testing business.' I tested stuff, things, product prototypes. Ideas would be the next big thing, the really big thing.

Being in the test business focused me on the outside, on the product. The philosophical notion of Inner versus and Outer Psychophysics was irrelevant during those days. It was product testing, focus on the external world, as ordered by the internal mind. There was no interest in mental processes, in so-called Inner Psychophysics. All the focus was on improving the product, keeping its cost down, keeping the acceptance at a level which allowed it to be purchased more than once by the consumer. The notion of mental processes and psychophysics applied to ideas, to inner constructs, was far away until a signal even occurred in 1980, one which would change my life forever.

As I wrote above, the notion of Inner constructs, ideas, was far from my mind. It was not relevant. Market researchers promoted their versions of concept testing, of segmentation, of a profound notion of the mind. Who was I to argue with these self-proclaimed experts of the mind? The notion of the mind had to be abandoned when I went into consumer research. Abandon, that is, until a call from Mr. Court Shepard, General Manager of Canada summoned me, and begin the unstoppable journal into the Inner Psychophysics. The story begins with Mr. Shepard asking me a very simple question, namely '*what do I say in order to sell more toothpaste*?' One could have picked me up from the ground. I was in shock. Here was a request from the top levels of Colgate Palmolive to answer the problem, a problem relevant to an iconic brand, Colgate Dental Cream.

Toothpaste, dental cream, sounds like a circuitous route by which Mind Genomics would be born, and perhaps Inner Psychophysics conceived, but that's the way it seems to me, at least now, as of this writing (Summer, 2018), looking back to June, 1980, 38 years before.

The approach was quite simple, reminding one of Norman Anderson's functional measurement (Anderson, 1970), albeit with a commercial twist, dealing with a product rather than an idea.

<u>Stage 1 – Design:</u>

1. We collected 90 or so different statements about dental cream in general, and about Colgate's product in particular.
2. We sorted these 90 statements into silos, groups, of similar meaning, such as '*how the product worked*,' '*the emotional benefit that using the product would bring*,' and so forth.
3. We then combined the elements into short, easy-to-read combinations, vignettes or test concepts, using an experimental design. We created 180 vignettes, twice as many as the 90 elements. Each element in a silo appeared an equal number of times.
4. The elements were statistically independent of each other.
5. Each vignette would have a minimum of two, and a maximum of five elements. This constraint made the vignette readable.
6. The vignettes were, therefore, incomplete by their very construction. That would prove no problem. Respondents were able to make their judgments with no apparent problem. In fact, it became obvious that respondents did not really 'read'

the vignette, but 'grazed' for information. Such grazing behavior is the norm, not the exception.

7. We simply ran the OLS on the data, to create a model showing the number of rating points on a 0-100 point scale would be contributed by each of our 90 elements.
8. We did this analysis several times, first with the total panel, then by key subgroup, such as age, gender, and brand used

Beyond dental cream – what was this really all about?

We might look at the foregoing as simply a questionnaire, requiring the respondent to take part in a survey about what is important in a dental cream. Looking at this method as simply one of the many types of surveys would not be particularly wrong, for it does, in fact, resemble a survey. The approach, systematically varying ideas, presenting them, getting respondents, and deconstructing the answers into the contribution of the components, first done on paper, later automatically by computer, seems patently simple. How could this possibly be an example of a method by which to implement this so-called 'Inner Psychophysics?' And indeed, for 35 years, three and a half decades, from 1980 to even after 2015, the notion was certainly not in our consciousness, that we doing a form of psychophysics. But we were. The psychophysics that we were doing was new to the world, an attempt to create a systematized way to metricize ideas. Casting aside the commercial rather than science origins of the effort, we might see in exercise the beginnings of an Inner Psychophysics. Our efforts were not directed towards uncovering relations between sensory magnitudes and physical magnitudes. Rather, we were metricizing thought, discovering the way the mind combines ideas. In so doing we were laying the groundwork to uncover the relations between ideas in the mind, doing so in the spirit of the Outer Psychophysics, but with ideas alone. We were creating a new discipline, Mind Genomics, with the promise of metricizing ideas in the mind, the same promise of a century before, but which had focused on metricizing sensory perception.

References

Anderson, N. H. Functional measurement and psychophysical judgment. Psychological Review, 1970, 77, 153-170.

Ekman, G., & Akesson, C. Saltness, sweetness and preference: A study of quantitative relations in individual subjects. Scandinavian Journal of Psychology, 1965, 6, 241-253.

Engel, R. Experimentelle Untersuchungen Uber die Abhangigheit der Lust und Unlust Von der Reizstarke beim Geschmacksinn. Pfluegers Archiv fur die Gesamte Physiologie, 1928, 64, 1-36.

Moskowitz, H.R., Kumaraiah, V., Sharma, K.N., Jacobs, H.L., & Sharma, S.D. Cross-cultural differences in simple taste preferences, Science, 1975, 190, 1217-1218.

CHILDREN'S EARLY NUMERICAL COMPETENCY AND PLAYWISELY

Craig Leth-Steensen
Carleton University, Ottawa, Canada
Patricia A. Hannan
PlayWisely, Dallas, TX, USA
Craig.LethSteensen@carleton.ca, hannanpatty3@gmail.com

Abstract

PlayWisely is an early learning program targeting basic cognitive and motor abilities in very young children. Key aspects of this program are also specifically designed to enhance school readiness, particularly with respect to reading and arithmetic. In this paper, the set of best early numerical competency predictors of children's mathematical ability in kindergarten (from Geary & vanMarle, 2016) is revealed and related to the math literacy training being provided by PlayWisely. Some discussion of current thinking regarding the manner in which children obtain cardinality proficiency is also discussed and related to PlayWisely's math literacy training.

"Children who score poorly on mathematics achievement tests at school-entry are at elevated risk for poor long-term mathematics outcomes" (Geary & vanMarle, 2016, p. 2130). Hence, it becomes extremely important to identify "the most critical early quantitative abilities that predict at risk children's later mathematics achievement" (Geary & vanMarle, 2016, p. 2132) in order to design the most appropriate types of remedial preschool interventions. In this light, the focus of the current paper is to highlight the fact that just such an intervention already exists, namely, the PlayWisely approach to early learning.

The PlayWisely Math-Literacy Card Sets

PlayWisely is the first comprehensive early learning method to include a multitude of fun activities specifically designed to foster both cognitive and motor skills starting from a very young age (i.e., 6 months of age and up). This innovative psycho-educational approach is designed to kickstart the mental processes responsible for attention, perception, reading and math, and the employment of both fine and gross motor skills by engaging children in playful activities that instruct and train the brain. PlayWisely involves both cognitive and physical/motor skills components. For the cognitive component, a flashcard system is used that is specifically designed to stimulate the development of key cognitive processing abilities and knowledge structures. The cognitive skills targeted by different sets of cards in this system include visual perception and attention (e.g., signal detection, acuity, visual-spatial tracking, and speed of processing), discrimination of key physical-based attributes (i.e., shape, colour, shading, size, and part-whole relationships), concept acquisition (i.e., classification and object recognition), and multi-sensory information integration (i.e., across the auditory and visual modalities). The card sets are also designed to develop foundational cognitive abilities associated with future numerical and literacy skills (e.g., subitizing, counting, learning the alphabet, phonemic awareness, and word recognition), hence, helping to ensure high levels of school readiness.

With respect to the issue of concern in this paper, it is the two Playwisely math literacy card sets that are most relevant (although related to these – but not specifically discussed here – are some other pre-math sets that are focussed on very early quantitative abilities). The math literacy sets include an initial set (called the #/# set) that consists of 10 cards with the numerical digits 1 through 10 along with 10 other cards showing the corresponding actual numerical quantities (where each week the digits are presented in different fonts and colors and the quantities are represented by a different animal face/small animal presented in organized dice-like patterns). The digit and quantity cards are presented twice through in an alternating fashion (e.g., “1”, “one dog”, “2”, “two dogs”, etc.; with the added name “dog/dogs” dropped the second time around) and then once more in a backwards fashion (just as for the pre-math sets described previously). For the younger children (less than 2 years old), typically only the digit and quantity cards for the numbers 1 through 6 are shown. An expansion of this technique is to alternate either the pitch (low and high, respectively) or the loudness (soft and loud, respectively) of the voice while presenting the digits and quantities. For older children (i.e., 2+) and children with a lot of experience with this set already, a key further expansion of this technique can also be employed. For this expansion, the digits 1 to 6 are lined up in order in a row on the floor and the children are asked to match up the corresponding quantities and place them in a row underneath the digits. Once all of the digits and quantities are in place, the children can be queried about the resulting rows and columns (e.g., “What’s in this row – numbers or dogs?” or “What’s the same about the things in this column?”). In addition to providing extensive repetition of the numerical sequence of single digits, both verbally and visually, this set also serves to introduce and then reinforce the associations that exist between numerical quantities and their corresponding numerical symbols.

The second math literacy set (called the +/- set) can be employed with older children (i.e., 2+) and children who can verbally count both forwards and backwards up to 6. This set begins with one image (e.g., a cat face, spider, or soccer ball with a different image each week) positioned at the top left of the first card and then incrementally adds more of the same image to this set one-by-one up to an eventual set of six (arranged in three rows of two on the final card). Each incremented image is first shown by itself on a transition card and then again in the same position as part of the incremented set. The first image and each subsequent incremented image are given names as they are presented (e.g., “This cat is for Owen, and this cat is for Felix, that makes two cats”, etc.). After incrementing up to the final card of six images, the set is incremented back down in reverse, taking away images one-by-one, until none are left (i.e., “All gone!”). One expansion of this technique as children become more experienced with this set over sessions (and begin to verbalize the correct quantities as they are incremented/decremented) is to drop the naming aspect and to start to use more standard arithmetic terminology (i.e.,” plus”, “minus”, and “equals”). This set is designed to help develop place-counting skills and to highlight the fact that each successive number in the count list is associated with the addition of one new element to the set of associated elements (which should also serve to stimulate early recognition of addition and subtraction concepts).

Which Are the Most Critical Early Abilities for Later Mathematics Achievement?

To answer this question, the findings reported in a recent study by Geary and vanMarle (2016) are extremely relevant. In this study, which is absolutely the most comprehensive one available to date, these authors collected initial measures of domain general abilities (i.e., executive functioning, verbal and non-verbal IQ, and letter recognition), as well as measures of both non-symbolic quantitative competencies (approximate number system [ANS] and approximate magnitude system [AMS] acuities, ordinal choice [between 1vs2 up to 6vs7 objects], object tracking system skill [involving only 2 objects or less], and non-verbal calculation) and symbolic quantitative competencies (enumeration, verbal counting, visual numeral recognition, and knowledge of cardinality). These measures were collected when the children ($N = 232$) in their study were an average of 3 years and 10 months old and used to predict scores on the Test of Early Mathematical Ability administered when these same children were an average of 5 years and 4 months old.

Three of the best unique predictors in this study (see Table 1) turned out to be the extent to which children could produce the sequence of numbers (i.e., verbal counting), could name visually presented digits (i.e., numeral recognition), and could provide exact quantities of objects associated with the digits 3 or more (i.e., had some knowledge of cardinality). Indeed, the key recommendation made by Geary and vanMarle (2016) was that these three symbolic-based competencies "may be more critical and foundational than others, and that these skills should be prioritized, especially for at-risk children" (p. 2139). At this point, it is important to note that all of these three skills are integral to the numerical training being provided by the PlayWisely math literacy #/# set.

Interestingly, in this same study, performance on acuity tasks involving either numerosity or physical size discrimination which are typically used to index non-symbolic-based quantitative competencies (assumed to be related to the ANS and the AMS, respectively) did not turn out to represent uniquely significant predictors of math achievement at the start of kindergarten (although each of these non-symbolic predictors was significantly related to mathematics achievement when correlated with this outcome on their own). Such findings are consistent with results obtained fairly recently by other researchers for slightly older (i.e., 6-8 years old) samples of children, who found that more exact, symbolic processing competencies (e.g., digit comparison) seem to win out over more approximate non-symbolic ones as predictors of mathematical ability (Sasanguie, Gobel, Moll, Smets, & Reynvoet 2013; see also Ostergran & Traff, 2013). Nonetheless, note that it is still possible that non-symbolic numerical knowledge "may contribute to some aspects of children's emerging understanding of symbolic mathematics, such as their cardinal knowledge" (Geary & vanMarle, 2016, p. 2138). Presumably, this contribution could involve learning to map core non-symbolic, analog-based numerosity representations to the numbers on a procedurally learned count list of integers (Wong, Ho, & Tang, 2016; where note that this is exactly the kind of training being provided by the PlayWisely #/# set). If so, such a contribution would then imply that the relation between non-symbolic numerical knowledge and mathematical outcomes is, in fact, mediated by children's corresponding symbolic-based knowledge (e.g., see Szkudlarek & Brannon, 2017, for some examples of such mediation).

Table 1. Final Regression Model Predicting Mathematical Achievement at Age 5 (Geary & vanMarle, 2016).

Initial Variable at Age 3:	*β*	*p*
Age (38 – 52 months)	.104	.0615
Girls vs. Boys	-.148	.1637
No information vs. College	-.124	.4300
Parents High School vs. College	-.111	.3802
Non-Verbal IQ	**.152**	**.0086**
Verbal Counting	**.182**	**.0055**
Numeral recognition	**.302**	**.0001**
One-knower vs. CP knower	**-.621**	**.0022**
Two-knower vs. CP knower	**-.372**	**.0448**
Three-knower vs. CP knower	-.203	.3245
Four-knower vs. CP knower	-.205	.2875
Ordinal Choice	**.120**	**.0266**

Note 1: All variables in bold are significant.
Note 2: All variables not included in this final model were already eliminated on the basis of a prior Bayes factor analysis (which allows for a determination of the importance of predictors that is less susceptible to the adverse effects of multicollinearity).
Note 3. Letter recognition was not included because it was highly collinear with numeral recognition.
Note 4. The test for cardinality knowledge was a "give-a-number" task which is scored categorically as the highest set-size for which children can give a correct number of items when queried (2 out of 3 times). Children who can do this for the number 5 or more are referred to as cardinal principle (i.e., CP) knowers.

Alternative Potential Routes to Cardinality Proficiency

Two other possible routes to cardinality proficiency (or full number-referent awareness), though, are thought to involve either enumerative counting (whereby the count word associated with the last thing counted represents the number of things in the set; i.e., often referred to as the cardinal principle, Sarneka & Carey, 2008, see also Leslie, Gelman, & Gallistel, 2008, and LeCorre, 2014) or subitizing (whereby the number of things in a small set can be discerned automatically as a whole) with the results of some experimental work by Benoit, Lehalle, and Jouen (2004) pointing towards subitizing as being the more likely route. In this light, it should be noted that the extent to which children could enumerate a set of items was also included in Geary and vanMarle's (2016) study but did not end up as a significant predictor in the final overall model (although, as for the non-symbolic-based acuity predictors, it was significantly related to mathematics achievement when correlated with this outcome on its own). Furthermore, it has also been shown to be the case that some children are able to procedurally count and respond correctly to an enumerated set of items when asked "how many" there are even though they lack full knowledge of the cardinality principle through other means of indexing it (e.g., they are classified as "subset-knowers" who can only provide specific quantities for numbers of 3-4 or less when asked to "give-a-number"; Sarnecka & Carey, 2008). Nonetheless, note that there is still a point in a further PlayWisely card set (i.e., the PlayWisely subject set) at which children are indeed asked to explicitly count out either 3 or 4 objects.

With respect to subitizing, one closely related view assumes that cardinal knowledge arises out of a representational system present in young infants (generally referred to as either the parallel individuation or object-tracking system; Le Corre &

Carey, 2007; Reynvoet & Sasanguie, 2016) that allows them to formulate mental models of small exact sets of individual items (i.e., of 3-4 items at most due to capacity limitations) that they then gradually learn to associate with the number names/digits on the count list (i.e., to subitize; where note that such associative knowledge would also likely be greatly facilitated by training on the PlayWisely #/# set). Full knowledge of cardinality for sets of items greater than 4 can then presumably occur by eventually noticing for these small known numerical sets that that the next number in the count list refers to the addition of one item (Le Corre & Carey, 2007; Reynvoet & Sasanguie, 2016).

Regardless of these different possibilities, one additional aspect of children's early number knowledge that is generally agreed upon to be extremely important (or even essential) is knowledge of the successor function (i.e., that "the cardinality for each numeral is generated by adding one to the cardinality for the previous numeral", Sarnecka & Carey, 2008, p. 665). Being able to apply semantic induction of this nature to the full number set (i.e., for both small and large numbers) typically occurs in close temporal proximity to the attainment of cardinality proficiency and is assumed to underlie a full command of the logic of natural numbers (Davidson, Eng, & Barner, 2012). In this light, the explicit training of this principle provided by the PlayWisely math literacy +/- set could be regarded as specifically targeting a key step in the development of numerical and, eventual, arithmetical proficiency.

Conclusion

It seems clear that the training being provided by the math literacy sets of the PlayWisely program is exactly of the kind that would be prescribed by current research in the area of early numerical/mathematical competency. Moreover, one key aspect of PlayWisely that can lead to particularly powerful and robust learning is the fact that children can start being administered the PlayWisely #-# set at very young ages (e.g., when they are less than 1 year old). Indeed, it seems fairly safe to conjecture that any child who has been repeatedly exposed to this set before the age of 3 will almost certainly be ahead of other children at Age 3½ or so with respect to all of the key predictive numerical competencies outlined here. This same child would then also be likely to benefit from the positive cascading effects of such early competency throughout the rest of his/her lifespan.

References

Benoit, L., Lehalle, H., & Jouen, F. (2004). Do young children acquire number words through subitizing or counting? *Cognitive Development*, *19*, 291-307.

Davidson, K., Eng, K., & Barner, D. (2012). Does learning to count involve a semantic induction? *Cognition*, *123*,162-173.

Geary, D. C., & vanMarle, K. (2016). Young children's core symbolic and nonsymbolic quantitative knowledge in the prediction of later mathematics achievement. *Developmental Psychology*, *52*, 2130-2144.

LeCorre, M. (2014). Children acquire the later-greater principle after the cardinal principle. *British Journal of Developmental Psychology*, *32*, 163-177.

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, *105*, 395-438.
Leslie, A. M., Gelman, R., & Gallistel, C. R. (2008). The generative basis of natural number concepts. *Trends in Cognitive Sciences*, *12*, 213-218.
Östergran, R., & Träff, U. (2013). Early number knowledge and cognitive ability affect early arithmetic ability. *Journal of Experimental Child Psychology*, *115*, 405-421.
Reynvoet, B., & Sasanguie, D. (2016). The symbol grounding problem revisited: A thorough evaluation of the ANS mapping account and the proposal of an alternative account based on symbol-symbol associations. *Frontiers in Psychology*, *7*, 1-11.
Sarneka, B. W., & Carey, S. (2008). How counting represents number: what children must learn and when they learn it. *Cognition*, *108*, 662-674.
Sasanguie, D., Gobel, S. M., Moll, K., & Smets, K., & Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number-space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology*, *114*, 418-431.
Szkudlarek, E., & Brannon, E. M. (2017). Does the approximate number system serve as a foundation for symbolic mathematics? *Language, Learning, and Development*, *13*, 171-190.
Wong, T. T., Ho, C. S., & Tang, J. (2016). The relation between ANS and symbolic arithmetic skills: The mediating role of number-numerosity mappings. *Contemporary Educational Psychology*, *46*, 208-217.

SESSION III

NEW MATHEMATICAL FORMALISM OF INNER PSYCHOPHYSICS

Ihor Lubashevsky
University of Aizu, Ikki-machi, Aizu-Wakamatsu, Fukushima 965-8580, Japan
i-lubash@u-aizu.ac.jp

First, Teghtsoonian's hypotheses (1971–1978) complemented by a number of results obtained later are discussed as arguments for the universality of inner psychophysics and the introduction of a psychological space as an individual object with its own properties. Then turning to the two-component description of human behavior (I. Lubashevsky, *Physics of the Human Mind*, Springer, 2017) the notion of mental space is formulated and human perception of external stimuli is treated as the emergence of the corresponding images in the mental space. On one hand, these images are caused by external stimuli and their magnitude bears the information about the intensity of the corresponding stimuli. On the other hand, the individual structure of such images as well as their subsistence after emergence is determined only by the properties of mental space on its own. Finally, the mental operations of image comparison and their scaling are defined in a way allowing for the bounded capacity of human cognition. Second, it is demonstrated, the developed theory of stimulus perception is able to explain the basic regularities of psychophysics, e.g., (*i*) the regression and range effects leading to the overestimation of weak stimuli and the underestimation of strong stimuli, (*ii*) scalar variability (Weber's and Ekman' laws), and (*iii*) the sequential (memory) effects. Finally, a solution to the Fechner-Stevens dilemma is proposed. This solution posits that Fechner's logarithmic law is not a consequence of Weber's law but stems from the interplay of uncertainty in evaluating stimulus intensities and the multi-step scaling required to overcome the stimulus incommensurability. Figure 1 illustrates this result.

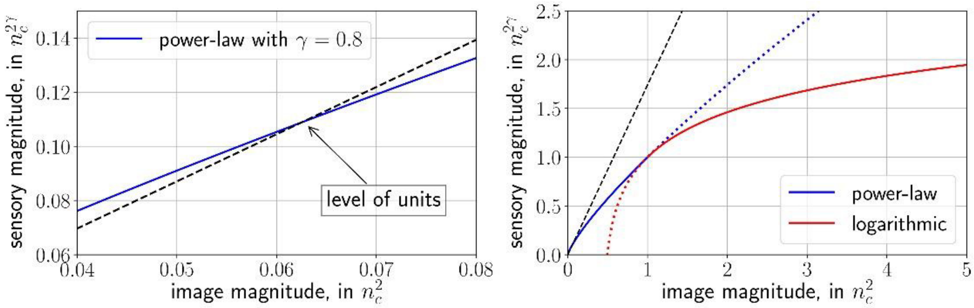


Fig. 1. Illustration of the relationship between the sensory magnitude and image magnitude constructed based on the quantitative comparison of images. Here dashed lines represent the image magnitude and in plotting $n_c = 4$ and $\gamma = 0.8$ were used. The quantity n_c^2 specifies the dynamic range of perceived magnitudes, the image magnitude is assumed to be determined by the intensity of neural signal generated by the corresponding sense organ. The introduction of the *level of units* implies that our life in the environment filled with various permanent stimuli should cause the emergence of their images in the mental space whose magnitudes are tacitly treated as the units in comparison with other images.

MODELING NEURAL RESPONSE CHANGE THRESHOLDS AS RANDOM VARIABLES WITHOUT MEMORY: IMPLICATIONS FOR PSYCHOPHYSICS

Wm Wren Stine
University of New Hampshire

Abstract

To construct a mathematical model of behavior measured psychophysically that purports to represent the behavior of the underlying neurophysiology, one often has to model the response rate of individual neurons. In such models, a given cell may be responding above the 'resting' rate when circumstances elicit a threshold change of response rate. The relevant question becomes how to characterize the probability of such a change. If one asserts that the probability of a threshold change is constant as a function of response rate, one has a stochastic process that has no memory, i.e., $P(X > x_B + x_T \mid X > x_B) = P(X > x_T)$*, where X is the continuous random variable representing the instantaneous response rate,* x_B *is the background response rate before the change, and* x_T *is the threshold response-rate change. X will thus have an exponential distribution, the implications of which will be explored.*

Models of perception, and behavior in general, that describe the neuroscience underlying behavior often fall on a continuum with models positing local representations by single cells, or units, forming one pole and models positing representations distributed across many units forming the opposite pole. Barlow's (1972; updated in 2009) Neuron Doctrine is an example outlining the local representation approach while Rumelhart, McClelland, & the PDP Research Group's (1986) Parallel Distributed Processing models are examples of the distributed representation approach. Specifically, "With a local representation, activity in individual units can be interpreted directly ... with distributed coding individual units cannot be interpreted without knowing the state of other units in the network" (Thorpe, 1995, p. 550; see also Page, 2000). One can consider a collection of units that forms a representation that may contain just one unit or many units depending on the model. Such a representing collection, of course, receives input from other collections and, in turn, communicates the results of its computation to other collections. Of course, the rate or pattern of response across the representing collection of units or cells codes the representation.

Without the loss of generality and within the context of a particular model, imagine a single unit, or cell, responding at a given rate. A change in that rate indicates a change in the representation that is to be communicated to other collections. If the change is too small, the collection receiving the communication will not change its response, or representation; the change in input is below threshold. So, we would have a change of input into the representing collection that changes its response accordingly and communicates that change to the receiving collection. Of critical interest is the probability that the change is communicated. One might expect that the probability of communicating a change could vary with the rate of response of the representing unit;

higher at a lower rate of response and lower at a higher rate of response, for example. If so, then the system's ability to switch representations will vary with the representations. Under what conditions will the probability of communicating a change remain constant as a function of the rate of response of the representing unit?

To answer that question, let where X be a continuous random variable representing the instantaneous response rate of the unit in question, x_B be the background response rate before the change in representation, and x_T be the threshold response-rate change in order to create a change in the response to which the collection is communicating (Filley, Khutoryansky, Dobias, & Stine, 2011, Appendix C.1). The probability that the response rate, X, exceeds threshold, x_T, is the same as that of exceeding some background response rate that is above threshold, $x_T + x_B$. That is,

$$P(X > x_B + x_T \mid X > x_B) = P(X > x_T). \quad (1)$$

So, for example, if the threshold response rate to create a change in response to the receiving units is $x_T = 10$ impulses per second, and the unit is responding at $x_B = 15$ imp/s, Equation (1) becomes

$$P(X > 25 \mid X > 15) = P(X > 10). \quad (2)$$

Notice that this condition does not imply that the probability that the response rate, X, exceeds threshold, x_T, is independent of the background response rate, x_B. For this example, independence would imply

$$P(X > x_B + x_T \mid X > x_B) = P(X > x_B + x_T)$$
$$\Rightarrow P(X > 25 \mid X > 15) = P(X > 25). \quad (3)$$

A random variable X that satisfies Equation (1) is said to represent a stochastic process with *no memory*, and X will follow an exponential probability density function (e.g., Meyer, 1970, p. 191; Weisstein, 2008),

$$f(x) = e^{-\lambda x}\lambda, \quad (4)$$

with cumulative distribution

$$f(x) = 1 - e^{-\lambda x}, \quad (5)$$

and a mean and standard deviation of $\frac{1}{\lambda}$. The difference between two independently- and identically-distributed exponential random variables follows a Laplace distribution while the sum of k independently- and identically-distributed exponential random variables follows an Erlang distribution.

The degree to which neurons have no memory, in the sense outlined previously, and the degree to which models of behavior posit neurons with no memory, will be degree to which such collections of neurons, or mathematical models of neurons, might be able to switch representations independently of the rate of response of the representing unit, or collection of cells.

Acknowledgements

Support in the form of both a Faculty Scholars Program Fellowship and a College of Liberal Arts sabbatical from the University of New Hampshire is gratefully acknowledged.

References

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology. *Perception*, 1, 371-394.

Barlow, H. B. (2009). Author's update: Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, 38, 795-807.

Filley, E. T., Khutoryansky, N., Dobias, J. J., & Stine, W. W. (2011). An investigation of the Venetian blind effect. *Seeing and Perceiving*, 24, 241-292.

Page, M. (2000). Connectionist modeling in psychology: A localist manifesto. *Behavioral and Brain Sciences*, 23, 443-512.

Meyer, P. L. (1970). *Introductory probability and statistical applications*, 2nd ed. Addison-Wesley.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (Eds.) (1986). *Parallel distributed processing: Explorations in the microstructure of cognition.* MIT Press.

Thorpe, S. (1995). Localized versus distributed representations. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (pp. 549-552). MIT Press.

Weisstein, E. W. (2008a). *Exponential distribution*, retrieved from MathWorld — A Wolfram Web Resource. Available at http://mathworld.wolfram.com/ExponentialDistribution.html.

ON THE ORIGINS OF THE POWER FUNCTION EXPONENT

Willy Wong
Dept. of Electrical and Computer Engineering and Institute of Biomaterials and Biomedical Engineering, University of Toronto
willy.wong@utoronto.ca

The exponent figures a prominent role in the power law proposed by Stevens and yet we know little of its origins. I will show that the exponent comes from the mathematical properties of a recently proposed class of probability distributions known as the Tweedie distribution. The Tweedie distribution has found application in modelling complex systems which exhibit long tail behaviour and overdispersion. The Tweedie family of distributions includes many of the common used probability distributions including the Gaussian, the Poisson and the gamma distributions. A key characteristic of a Tweedie distribution is the interrelation between its first and second statistical moments. Using the entropic hypothesis of Ken Norwich, I will show that the Tweedie distribution provides a first theoretical understanding of the power function exponent, and its connection to the many properties of sensory transduction and psychophysics.

MATHEMATICAL DEFINITION OF EVIDENCE IN BAYESIAN MODEL OF DECISION MAKING AND CONFIDENCE ESTIMATION IN SENSORY TASKS

Valentin M. Shendyapin
Institute of Psychology Russian Academy of Sciences. Moscow. Russia
valshend@yandex.ru

Abstract

Several models of decision making in sensory discrimination use the notion of evidence in favor of a certain response hypothesis. But mathematical questions were not discussed in these models about the evidence forming and its relations with correctness probabilities of hypotheses compared. The model suggested tries to solve these problems by involving a new definition of the evidence notion into Signal Detection Theory. It was shown, firstly, that the correctness probabilities of the hypotheses are monotonic hyperbolic tangent functions of the defined evidence itself and, secondary, that resulting evidence is a sum of independent partial evidences. These properties of the evidence totally correspond to our intuitive understanding of evidence notion, usable by a judge for decision making.

To start confidence modeling we have to answer a general question first: what is a function of confidence as a mental phenomenon? Now the concept of confidence has no accurate and unambiguous definition either in technical sciences, or in humanities. In the case of psychophysics, confidence isn't defined properly not only for a real observer, but even for an ideal one. Since the behavior of an ideal observer is much easier to explain, than that of a real one, then the consideration of an ideal observer's confidence is the first thing we have to do.

At the same time we believe that both the ideal observer, and the real one reflects the outside world and own internal state correctly in general, and also control his (her) own activity successfully. The behavior of the real observer may be less efficient than that of the ideal one, but it cannot differ from rational behavior cardinally.

Along with direct experimental observations of brain activity, researchers look for theoretical computational schemes on the basis of which brain's neurons provide decision making when performing sensory tasks. Signal Detection Theory (SDT), which is based on Bayesian inference, is used often in such work. SDT searches for statistical decisions (i.e. decision variables and decision making rules) for determination of an unknown parameter of statistical hypotheses. These hypotheses are formulated according to a subject's task for available successive samples of an input random variable, which does not coincide with a decision variable as a rule. The feature of SDT: such the input variable is a sensory effect of a stimulus's parameter and two alternative hypotheses are distributions of this parameter for two different stimuli.

Any influence of an object observed on a sense organ is perceived by receptors which are sensitive to special stimuli varying frequency of output spikes.

Since all receptors work imperfectly, there are always random noises in their outputs along with useful signals. This sum of signal and noise passes through several layers of neuronets. As a result a sensory image of an object observed appears in the brain where output signals of all sensory systems come. A random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ is used commonly as the simplest mathematical representation of this image. Particular components $X_1, X_2, \ldots, X_n$ describe a form, color, a smell and other more complicated features of the object perceived. We shall use a vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ for designation of the numerical values of the vector $\boldsymbol{X}$ random components which have been received as a result of the observation. Because of noises in vector's components, a final decision may be erroneous.

Now we can give our answer to the question what is confidence. From the very beginning of our work we made up a hypothesis that confidence feeling is an efficiency indicator of a response chosen, which may be more easily calculated than the efficiency itself. And now let's find correct formulas for a confidence value calculation.

Mathematical model of 2 similar multi-dimensional stimuli discrimination

Our model describes the following experimental task. 2 similar stimuli (*A* and *B*) are presented randomly to an observer. The observer has: 1) to identify a stimulus presented by the response *a* (the stimulus *A* has been presented) or *b* (the stimulus *B* has been presented); 2) to estimate confidence in the decision made. Awards are given for correct responses: $v_{Aa} > 0$, $v_{Bb} > 0$ and penalties for erroneous ones:
$v_{Ab} < 0$, $v_{Ba} < 0$.

Suppose that we have a vector of numerical values of independent random stimulus parameters $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ which is obtained during the observation. Due to noises in sensory system the received vector $\boldsymbol{x}$ can correspond to both stimuli *A* and *B*. If a priory probabilities of stimuli *A*, *B* presentations $P(A)$, $P(B)$, and a priory distributions of probabilities $f(\boldsymbol{x}|A)$, $f(\boldsymbol{x}|B)$ are known for all possible $\boldsymbol{x}$, then according SDT (Egan, 1975), a posteriori probability of each stimulus $P(A|\boldsymbol{x})$, $P(B|\boldsymbol{x})$ and a posteriori average utility for each response $V(a|\boldsymbol{x})$, $V(b|\boldsymbol{x})$ may be calculated for this vector $\boldsymbol{x}$:

$$\psi(\boldsymbol{x}) = [P(A)\, f(\boldsymbol{x}|A)] \,/\, [P(B)\, f(\boldsymbol{x}|B)], \quad (1)$$
$$P(A|\boldsymbol{x}) = \psi(\boldsymbol{x}) \,/\, [1 + \psi(\boldsymbol{x})]\,, \quad (2)$$
$$P(B|\boldsymbol{x}) = 1 \,/\, [1 + \psi(\boldsymbol{x})]\,, \quad (3)$$
$$V(a|\boldsymbol{x}) = P(A|\boldsymbol{x})\, v_{Aa} + P(B|\boldsymbol{x})\, v_{Ba}\,, \quad (4)$$
$$V(b|\boldsymbol{x}) = P(A|\boldsymbol{x})\, v_{Ba} + P(B|\boldsymbol{x})\, v_{Bb}\,, \quad (5)$$

Using these formulas an experienced mathematician can calculate values of $P(A|\boldsymbol{x})$ and $P(B|\boldsymbol{x})$, $V(a|\boldsymbol{x})$ and $V(b|\boldsymbol{x})$. After comparing the values in the each pair, he (she) can choose a response easily. But brain's neurons of the every living being are not mathematicians; they can't make these complicated calculations. In addition, formulas (1 – 5) do not allow giving a definition for confidence. The main problem for multi-dimensional stimuli discrimination task is to give such a reasonable definition, which will be convenient for neuronal calculations.

There are several SDT-based models (Balakrishnan, Ratcliff, 1996; Bjorkman et al., 1993; Ferrel, 1995; Pleskac, Busemeyer, 2010), which describe both decision making and confidence estimation for one-dimensional stimuli discrimination. It is assumed that an observer uses a decision making criterion which is localized on the axis of the one-dimensional sensory effect x. After making a response choice, the observer estimates confidence as a distance between the criterion and the sensory effect x obtained during the observation. This way of confidence estimation may be applied to the "greater–lesser" task. However, it isn't applicable for the "same–different" task, because in this case there are two criteria on the sensory axis. In this case the observer does not have any instruction showing which of them should be used for the confidence assessment. As for multi-dimensional stimuli, this way to estimate confidence isn't applicable even for the "greater–lesser" task, because we have many sensory axes. What axis we have to use in this case?

Random Walk Models (Link, Heath, 1975; Heath, 1984) and Accumulator Model (Vickers, Lee, 1998) have avoided this problem by assuming that brain's neurons convert the sensory effect into evidences favoring each response alternative. These evidences are integrated over time and a decision is made when a sufficient evidence has been accumulated which favor one alternative over another. Many models, for example, Two-Alternative Forced-Choice (TAFC) task models and Drift Diffusion Model (DDM), also use the notion of evidence favoring each alternative in decision making process (Bogacz et al., 2006). But up to the present day, questions were not discussed in mathematical psychophysics about the evidences forming and their relations to correctness probabilities of hypotheses compared.

We made an assumption that for multi-dimensional discrimination it is necessary to integrate all sensory effects $x_1, x_2, \dots, x_n$ into one universal decision variable, which permit both to choose response and to estimate confidence. The model suggested (Shendyapin, 2016) has solved the problems mentioned by involving a new definition of the evidence in favor of the stimulus A choice ($\Psi_A(\boldsymbol{x})$) into SDT:

$$\Psi_A(\boldsymbol{x}) = ln\{[P(A)f(\boldsymbol{x}|A)]\,/\,[P(B)f(\boldsymbol{x}|B)]\} = \Lambda_A + L_A(\boldsymbol{x}), \qquad (6)$$

where $\Lambda_A = \ln[P(A)/P(B)]$ — is a natural logarithm of a ratio of a priory probabilities of the stimuli A and B, which may be called a priory evidence in favor of the stimulus A choice; $L_A(\boldsymbol{x}) = \ln[f(\boldsymbol{x}|A)]\,/\,f(\boldsymbol{x}|B)]$ — is a natural logarithm of a likelihood ratio of the stimuli A and B, which may be called a posteriori evidence in favor of the stimulus A choice.

Formula (6) shows that the observer forms a scalar value of evidence $\Psi_A(\boldsymbol{x})$ on the base of a priori information of the task characteristics (constants $P(A)$, $P(B)$ and functions $f(\boldsymbol{x}|A)$, $f(\boldsymbol{x}|B)$) obtained previously (for example, by learning). But the numerical vector of the stimulus parameters: $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$, obtained during the current observation, is needed as an argument for functions $f(\boldsymbol{x}|A)]$, $f(\boldsymbol{x}|B)$ also.

If particular random components $X_1, X_2, \dots, X_n$ of the vector $\boldsymbol{X}$ are independent, then a posteriori evidence in favor of the stimulus A choice may be calculated as:

$$L_A(\boldsymbol{x}) = \ln[f(x_1|A)] - \ln[f(x_1|B)] + \ln[f(x_2|A)] - \ln[f(x_2|B)] + \dots$$
$$+ \ln[f(x_n|A)] - \ln[f(x_n|B)]. \qquad (7)$$

Thus the resulting evidence $\Psi_A(\boldsymbol{x})$ (6) is a sum of a priory evidence in favor of the stimulus A choice: $\Lambda_A = \ln[P(A)] - \ln[P(B)]$ and a posteriori independent partial evidences: $L_A(x_1) = \ln[f(x_1|A)] - \ln[f(x_1|B)]$, $L_A(x_2) = \ln[f(x_2|A)] - \ln[f(x_2|B)]$, … , $L_A(x_n) = \ln[f(x_n|A)] - \ln[f(x_n|B)]$.

Using the expression (6) for a scalar value of the evidence $\Psi_A(\boldsymbol{x})$, we have derived formulas for probabilities of the stimuli *A* and *B* identification correctness and for the average utilities of the responses *a* and *b*:

$$\Psi_A(\boldsymbol{x}) = \Lambda_A + L_A(\boldsymbol{x}) = \Lambda_A + L_A(x_1) + L_A(x_2) + \ldots + L_A(x_n). \quad (8)$$
$$P[A|\Psi_A(\boldsymbol{x})] = 0.5\{1 + th[\Psi_A(\boldsymbol{x})/2]\}, \quad (9)$$
$$P[B|\Psi_A(\boldsymbol{x})] = 0.5\{1 - th[\Psi_A(\boldsymbol{x})/2]\}, \quad (10)$$
$$V[a|\Psi_A(\boldsymbol{x})]= 0.5\{(v_{Aa} + v_{Ba}) + (v_{Aa} - v_{Ba})\ th[\Psi_A(\boldsymbol{x})/2]\}, \quad (11)$$
$$V[b|\Psi_A(\boldsymbol{x})] = 0.5\{(v_{Ab} + v_{Bb}) + (v_{Ab} - v_{Bb})\ th[\Psi_A(\boldsymbol{x})/2]\}, \quad (12)$$

where function $th(x)$ is hyperbolic tangent. If x increases, then $th(x)$ increases monotonically as well. It is well known that a typical reaction of a neuron seems to have a logarithmic form. The hyperbolic tangent seems as typical reaction of a neuron as well. Therefore all formulas derived (8 – 12) are more convenient for neuronal calculations, than known SDT formulas (1 – 5).

Obviously, the evidence $\Psi_A(\boldsymbol{x})$ (6) may be used as the decision variable for a usual discrimination task. When the variable $\Psi_A(\boldsymbol{x})$ increases, then the probability $P[A|\Psi_A(\boldsymbol{x})]$ and the average utility $V[a|\Psi_A(\boldsymbol{x})]$ increase monotonically as well. At the same time $P[B|\Psi_A(\boldsymbol{x})]$ and $V[b|\Psi_A(\boldsymbol{x})]$ decrease monotonically. Therefore $\Psi_A(\boldsymbol{x})$ is "the evidence in favor of the stimulus *A*" indeed. The pairs of functions $P[A|\Psi_A(\boldsymbol{x})]$, $P[B|\Psi_A(\boldsymbol{x})]$ and $V[a|\Psi_A(\boldsymbol{x})]$, $V[b|\Psi_A(\boldsymbol{x})]$ have only one intercept on the axis $\Psi_A(\boldsymbol{x})$. Therefore the ideal observer always has only one decision making and confidence estimation criterion $\Psi_{A,\mathrm{cr}}$ on the axis $\Psi_A(\boldsymbol{x})$.

Thus by means of SDT development, we have suggested a new approach for modeling of decision making and confidence estimation in discrimination tasks. This approach is rather general since it is suitable for arbitrary distributions $f(\boldsymbol{x}|A)$, and $f(\boldsymbol{x}|B)$. Thus it is suitable both for the "greater–lesser" sensory task and for the "same–different" one.

In order to estimate efficiency (a correctness probability or an average utility) of the response chosen, we may to define confidence as a distance between the criterion $\Psi_{A,\mathrm{cr}}$ and the numerical value of the evidence $\Psi_A(\boldsymbol{x})$ obtained as a result of the observation: $C = \Psi_A(\boldsymbol{x}) - \Psi_{A,\mathrm{cr}}$. If the ideal observer's goal is to choose a response, which has the greatest correctness probability, then $\Psi_{A,\mathrm{cr}} = 0$ and confidence is $C_{\mathrm{cor}}(\boldsymbol{x}) = \Psi_A(\boldsymbol{x}) = \Lambda_A + L_A(\boldsymbol{x})$. An observer's response is *a*, if $C_{\mathrm{cor}}(\boldsymbol{x}) > 0$, otherwise it is *b*.

If the goal is to choose a response, which has the greatest average utility, then $\Psi_{A,\mathrm{cr}} = -\Sigma$, where $\Sigma = ln[(v_{Aa} - v_{Ab})/(v_{Bb} - v_{Ba})]$, and confidence is $C_{\mathrm{util}}(\boldsymbol{x}) = \Psi_A(\boldsymbol{x}) + \Sigma = \Lambda_A + L_A(\boldsymbol{x}) + \Sigma$. The observer's response is *a*, if $C_{\mathrm{util}}(\boldsymbol{x}) > 0$, otherwise it is *b*. Thus, confidence feeling may be used for both the decision-making and the probabilistic evaluation of its efficiency.

Acknowledgements

The work was supported by Russian Foundation of Basic Research. Project № 18-013-00148. The author thanks Irina Skotnikova for her help and very useful discussions of the model psychological content.

References

Balakrishnan, J.D., Ratcliff, R. (1996). Testing models of decision making using confidence ratings in classification. *Journal of Experimental Psychology: Human Perception & Performance,* 22, 615–633.

Bjorkman, M., Juslin, P., Winman, A. (1993). Realism of confidence in sensory discrimination: The underconfidence phenomenon. *Perception & Psychophysics,* 54, 75–81.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P. and Cohen, J.D. (2006). The Physics of Optimal Decision Making: A Formal Analysis of Models of Performance in Two-Alternative Forced-Choice Tasks. *Psychological Review*, 113(4), 700–765.

Egan, J.P. (1975). *Signal Detection Theory and ROC Analysis.* N.Y.: Academic Press.

Ferrel, W.R. (1995). A model for realism of confidence judgments: implications of under confidence in sensory discrimination. *Perception & Psychophysics,* 57, 246–254.

Heath, R.A. (1984). Random-walk and accumulator models of psychophysical discrimination: a critical evaluation. *Perception,* 13, 57–65.

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

Pleskac, T.R., Busemeyer, J.R. (2010). Two-Stage Dynamic Signal Detection: A Theory of Choice, Decision Time, and Confidence. *Psychological Review,* 117(3), 864–901.

Shendyapin, V.M. (2016). Development of confidence model on the basis of Signal Detection Theory. *Fechner Day, 2016. Proc. 32 Ann. Meet. ISP* (p. 33). Moscow. Russia.

Vickers, D., Lee, M. (1998). Dynamic models of simple judgments: I. Properties of a self-regulating accumulator model. *Nonlinear dynamics, psychology and life sciences.* 2, 169–194.

SESSION IV

CRITICAL THINKING IN THE MONOLINGUAL AND BILINGUAL POPULATION

Sophia Arndt[1], Shimeng Liu[2], Yuko Yamashita[3], Yoshitaka Nakajima[2], Stanislava Antonijevic-Elliott[1], and Mark A. Elliott[1]
[1]*National University of Ireland, Galway*
[2]*Department of Human Science/Research Center for Applied Perceptual Science, Faculty of Design, Kyushu University, Fukuoka, Japan*
[3]*Shibaura Institute of Technology, Tokyo, Japan*
s.arndt2@nuigalway.ie

This study aims to examine the relationship between bilingualism and critical thinking. Monolingual and bilingual speakers completed the following online questionnaires: Participants were assessed on their language experience and proficiency (LEAP-Q), followed by a newly developed and validated critical thinking assessment. This evaluated their ability in five dimensions of critical thinking; hypotheses testing, verbal reasoning, judging likelihood, argumentation analysis and problem-solving, following Halpern's (2002) construct definitions. Outcomes of within-group comparisons were found to be noteworthy. Parallels were observed in all languages, showing an advantage of solving verbal-reasoning tasks in the first language. Questions on judging-likelihood seem to be solved better in an individuals' second language.

Findings of this research cover cross-language differences and allow insight into the possible influence of cross-cultural differences. Outcomes can be attributed to fundamental cognitive performance rather than differences between specific languages or cultural contexts.

References

Halpern, D. F. (2002). *Thought and knowledge: An introduction to critical thinking*. Routledge.

Marian, V., Blumenfeld, H. K., & 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), 940-967.

A SPECTRAL-CHANGE FACTOR RELATED TO SONORITY MAKES NOISE-VOCODED JAPANESE SPEECH INTELLIGIBLE

Takuya Kishida[1], Yoshitaka Nakajima[2], Kazuo Ueda[2], Gerard B. Remijn[2], and Seiya Umemoto[3]

[1] *Department of Human Science, Kyushu University, 4-9-1 Shiobaru Minami-ku Fukuoka, Japan*

[2] *Department of Human Science,/Research Center for Applied Perceptual Science, Kyushu University, 4-9-1 Shiobaru Minami-ku Fukuoka, Japan*

[3] *21st Century Program, Kyushu University, 744 Motooka Nishi-ku, Fukuoka, Japan*

kishida.takuya0119@gmail.com, nakajima@design.kyushu-u.ac.jp, ueda@design.kyushu-u.ac.jp, remijn@design.kyushu-u.ac.jp, 1NC15002K@s.kyushu-u.ac.jp

Essential cues to perceive speech were clarified by measuring intelligibility of noise-vocoded Japanese speech resynthesized from 2–4 spectral-change factors. Perceptual cues in spectral change of speech sounds are highly redundant, thus we summed up the spectral change into small number of factors by using a statistical method: origin-shifted principal component analysis followed by varimax rotation. Our previous studies found that intelligible noise-vocoded Japanese speech can be resynthesized when 3 or more factors are used (Kishida et al., 2016). We further found that when the number of factors was 3 or 4, a factor with a high factor loading appeared in a range around 500–1500 Hz. The factor score distribution of this factor was highly correlated with *sonority* (Nakajima et al., 2017). We hypothesized that this sonority factor would reflect important cues for speech perception. In the present study, sixteen native speakers of Japanese listened to noise-vocoded Japanese speech and reported morae (syllable-like phonological units) which they heard. Spectral change of the noise-vocoded speech was reconstructed from 2, 3, and 4 factors selected from each available set of factors obtained from the 2-, 3- and 4-factor analyses. In the 2-factor condition, for example, there were 10 possible factor combinations; 6 combinations of factors from the 4-factor analysis, 3 combinations of factors from the 3-factor analysis, and 1 combination of factors from the 2-factor analysis. When the sonority factor was omitted from resynthesizing, lower mora identification was obtained than when this factor was used even if the total number of factors remained the same. If the number of factors was 2, however, mora identification was relatively low (around 30%) whether or not the sonority factor was used. The results indicate that 1) using 3 or more factors and 2) using the sonority factor are both required for synthesizing noise-vocoded speech of which intelligibility exceeds 50% from spectral-change factors.

References

Kishida, T., Nakajima, Y., Ueda, K., & Remijn, G. B. (2016). Three factors are critical in order to synthesize intelligible noise-vocoded Japanese speech. *Frontiers in psychology*, 7:517.

Nakajima, Y., Ueda, K., Fujimaru, S., Motomura, H., & Ohsaka, Y. (2017). English phonology and an acoustic language universal. *Scientific Reports*, 7:46049.

DO PROSODIC CUES NEED A VISA TO TRAVEL? THE PERCEPTION OF EMOTIONS IN SPEECH ACROSS LANGUAGES: EVIDENCE FROM HEBREW, ENGLISH AND GERMAN.

Vered Shakuf[1], Cilla Zack[1], Lior Tidhar[1], Shira Golani[1], Boaz M. Ben-David[1], Sarah Ben-David[2], Thomas Lachmann[3], Patricia de Brito Castilho Wesseling[3], Sabrina Defren[3], Thomas Wegner[3], and Shanley Allen[4]

[1] *Communication Aging and Neuropsychology (CANlab) lab, Baruch Ivcher School of Psychology, Interdisciplinary Center (IDC) Herzliya, Israel*

[2] *Department of Criminology, Ariel University, Ariel, Israel*

[3] *Cognitive and Developmental Psychology, University of Kaiserslautern, Germany*

[4] *Psycholinguistics and Language Development Group, University of Kaiserslautern, Germany*

boaz.ben.david@idc.ac.il

Abstract

The ability to correctly perceive emotions in speech stands at the core of human communication. To identify an emotion, one should be able to process the semantic (lexical meaning) and the prosody (tone of speech) of the utterance, and integrate them. While both universal and language-specific aspects of emotional speech perception are mentioned in the literature, to date, the relation between these factors remains vague. We examined the universality of linguistic and cultural factors in emotional speech using a novel tool, T-RES, Test for Rating of Emotions in Speech, adapted and validated to three languages: Hebrew, English and German. We present preliminary data from a set of studies highlighting surprising similarities across languages and cultures in prosodic cues processing and the rules for integrating speech channels. Another intriguing question tested is the issue of processing emotions in speech for non-native speakers. We compared perception of emotions in speech between native and non-native speakers of Hebrew and English. Preliminary findings suggest that listeners provide an even larger bias to the prosody (over semantics) in their L2 than in their L1, across English and Hebrew native speakers. In sum, we suggest that even though basic features of processing emotional speech remain the same across languages and cultures, there may be difference in the processing of emotions in a listener's L1 vs L2.

In the global world of the 21st century, cooperation and collaboration between different countries is an important part of almost every industrial, business and academic activity. Often, the success of this type of cooperation relies on participants' ability to communicate efficiently in their second language. This becomes an even more pressing issue when considering immigrants who are living in a new country, struggling to adjust to a new language and culture. Accurate recognition of emotion is a key element of successful communication. To take part effectively in social interaction, it is essential to identify, understand and respond appropriately to the emotion conveyed in speech. Emotions in speech are conveyed via two auditory channels: Semantics (the lexical content of the words) and prosody (i.e., rhythm, stress, and intonation of speech: Ben-David, Multani, Shakuf, Rudzicz, & van Lieshout, 2016). Accumulating evidence in the literature points to both universal, and language- and culture- specific principals governing the communication and recognition of emotions in speech (Pell, Paulmann, Dara, Alasseri, & Kotz, 2009). Cross-cultural (or cross-linguistic) differences in the perception of emotions in speech can be a result of using different prosodic cues and intonational patterns in different languages and cultures (Scherer, Banse, & Wallbott, 2001). It can also be a result of differences in the manner in which the two speech channels (i.e., semantics and prosody) are integrated (Ishii, Reyes, & Kitayama, 2003; Kitayama & Ishii, 2002). Due to these differences, it is reasonable to assume that deciphering the speaker's original emotional message would be even more challenging when communicating in one's second language (L2) rather than in the first (L1). However, to date, the relative role of each of these factors remains unclear.

Consider the following example: you receive a phone call from a colleague from an overseas university saying "I feel wonderful today" spoken with angry prosody. What will you think? Is he/she expressing happiness, anger, or a combination of the two? Phrasing the above question more generally, are listeners from different cultural (or linguistic) backgrounds able to perceive and correctly identify non-verbal emotional cues conveyed by their counterparts? Do they assign the same weights to the two speech channels (semantics and prosody) as intended by the speaker?

Universal vs. culture- and language-specific principles of emotion recognition in speech

Listeners, when asked, are able to identify emotions conveyed in speech relying only on prosodic information (Paulmann & Uskul, 2014). The evidence in the literature points to both universal and culture-specific principles for decoding emotional prosodic cues. When members of one cultural group were asked to identify emotions uttered by speakers from different cultural groups, accuracy was found to be better than chance across emotions and cultures. However, in all these studies listeners showed an in-group advantage for decoding emotional prosody. Namely, correct identification rates were higher for identifying emotions in their own language or spoken by a member of their cultural group (Scherer et al., 2001).

Differences in integration of speech channels

Contrary to the abundance of literature on the universality of prosodic cues, only a few studies have investigated culture-related differences in the integration of emotional speech channels (prosody and semantics). For example, Min and Schrimer (2011) presented native and non-native English speakers with emotional words that were spoken either in matching or mismatching prosody. Both groups responded faster and more accurately when semantics and prosody were congruent as compared to when they were incongruent, with no significant difference between the two groups. In contrast, in a series of studies, Ishii, Kitayama and their colleagues (2003; 2002) showed that it was more difficult for Western listeners to ignore the semantic domain (when asked to focus on the prosody) than to ignore the prosody (when asked to focus on the semantics). However, for members of Eastern cultures, the reverse effect appears. The Ishii and Kitayama team has related this divergence in attentional bias to cultural aspects, rather than linguistic ones.

Using a novel tool, T-RES, Test for Rating of Emotions in Speech, we examined the universality of linguistic and cultural factors in emotional speech. In a set of three studies, we addressed three main questions: (1) Do native speakers of different languages from different countries (specifically, Israelis native speakers of Hebrew and Canadians native speakers of English) process emotional speech in a different manner? (2) Are listeners from different countries (specifically, Germans and Israelis) able to perceive and correctly identify non-verbal emotional cues conveyed in the same language (Hebrew)? (3) Is emotional speech processed differently in a listener's first language (L1) versus in his/her second language (L2, specifically, English and Hebrew).

T-RES, Test for Rating of Emotions in Speech

This set of studies employ the existing Hebrew and English versions (Ben-David et al., 2016; Shakuf et al., 2016) of the T-RES, a novel test designed to expose the full complexity of the interaction between semantic and prosodic channels in spoken emotions with validated linguistic properties (Ben-David et al., 2013, 2011). In the T-RES, listeners are presented with spoken sentences that carry emotions in the semantics, prosody or both. In each block, participants are asked to rate each spoken sentence on four different 6-point Likert scales, relating to how much they agree that the speaker expressed a specific emotion (e.g., anger). In the prosody- and semantic-rating tasks, listeners are asked to selectively attend only to one channel, ignoring the other. In the general-rating task, listeners rate the sentence as a whole, using the information in both prosodic and semantic channels. Four affective categories are tested: anger, fear, happiness, and sadness. A neutral category was added as a baseline condition for performance.

Study 1 - Comparing Hebrew and English

Study 1 evaluates differences and similarities between-Hebrew-speaking Israelis and English-speaking Canadians in the processing of both speech dimensions in tandem. To that end, we used the English version of the T-RES to test Canadian native English speakers and the Hebrew version of the T-RES to test Israeli native Hebrew speakers.

Preliminary results provide evidence for the following universal principles in the integration of emotional speech channels in English and Hebrew: *(1) Congruency supremacy* – Both Canadian and Israeli listeners (in respective languages) rated sentences that present the same emotion in both channels as presenting a larger extent of emotion (on redundancy see, Ben-David & Algom, 2009; Ben-David, Eidels, & Donkin, 2014; *(2) Prosodic Dominance* - when rating an emotion in speech, prosodic information plays a larger role than semantics for both native Hebrew and native English speakers (in respective languages). For an example of the results from the anger-rating task, see Figure 1.

Study 2 - Universality of Prosodic cues

Study 2 was designed to test the universality of prosodic cues across German and Hebrew. Specifically, we tested whether listeners can identify emotions expressed by prosodic cues in a foreign language, without access to the semantic content. We used sentences taken from the Hebrew version of the T-RES to test native German-speaking German students that do not speak Hebrew. Their performance was compared to native Hebrew-speaking Israeli students.

These two languages are used as they represent languages that come from different language families – German is an Indo-European language, whereas Hebrew is a Semitic language, yet they are both phoneme-based languages. Hebrew and German have a very similar prosodic structure but differ in specific suprasegmental phonological properties.

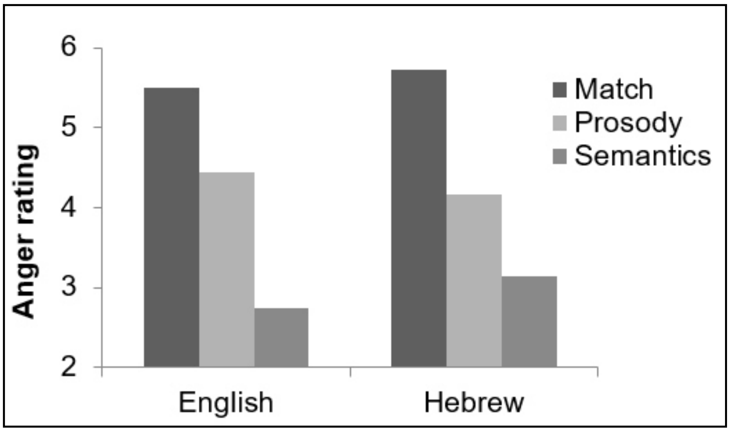


Figure 1. Anger ratings of spoken sentences that present anger, either in both prosody and semantics (Match), only in the prosody or only in the semantics. Left bars - English sentences rated by Canadian native English speakers. Right bars - Hebrew sentences rated by Israeli native Hebrew. The figure indicates that in both languages the same trends appears match > prosody > semantics.

The two also represent different cultural backgrounds, which may have significant impacts on the way listeners process emotions. To the best of our knowledge no direct comparison of Hebrew and German has been performed.

Our preliminary findings suggest that German listeners were able to identify prosodic emotions spoken in Hebrew by a Hebrew speaker, even though they did not have access to the semantics of the speech. This trend hints on the possibility of universal principles in decoding emotional prosodic cues, as listeners can correctly rate emotional prosody in a foreign unknown language. Figure 2 illustrates that the pattern of misattribution of emotions was also highly similar across groups. When listeners were asked to focus on the prosody, both groups rated sentences spoken with sad prosody very highly on the sadness scale (5.8-5.6/6). When asked to rate these same prosodically sad sentences in terms of prosodic fear, again both groups provided highly similar ratings (3.1-3.8/6), rather than rating the sentences as not representing any fear (e.g., rating of ~1/6). However, neither group rated the same prosodically sad sentences as expressing even a modicum of anger or happiness (1.1-1.4/6). In other words, both Hebrew and German speakers ranked Hebrew sentences spoken with sad prosody as expressing some fear, but no anger or happiness.

Study 3 – Comparing Native Language (L2) vs Non-Native Language (L2)

In Study 3, we tested for similarities and differences in the perception of emotional speech between native (L1) and non-native (L2) speakers of English and Hebrew, comparing performance on the English and Hebrew versions of the T-RES. Two groups of participants were tested: (1) Israel residents, native English speakers, whose second language is Hebrew (L1-E & L2-H); (2) Israeli native Hebrew speakers, whose second language is English (L1-H & L2-E). The T-RES was administrated to each participant in both languages, always beginning with the participant's L1.

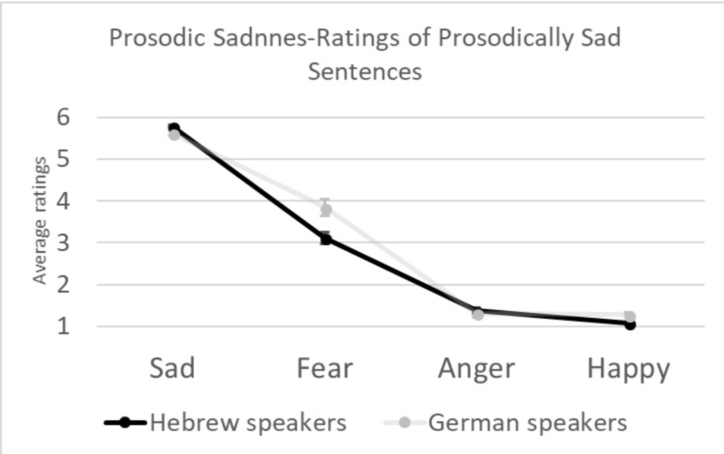


Figure 2. Average ratings of prosodic sadness for Hebrew sentences spoken with sad prosody by Native-Hebrew speaking Israeli students (black line) and German students who do not speak Hebrew (grey line).

Our preliminary findings highlight *Prosodic dominance* – when rating an emotion in speech, prosodic information plays a larger role than semantics in L1, and significantly more so in L2, for both native Hebrew and native English speakers. This is illustrated in Figure 3, where prosodic dominance is evident both in L1 and L2, but the interaction indicates that prosodic dominance is larger in L2, across participants (main panel) and separately for native Hebrew and native English speakers (side panels). These results suggest the unique processing of speech in a listener's second language, given differences in auditory processing in L2 (Ben-David, Avivi-Reich, & Schneider, 2016).

Discussion

Preliminary data collected from three studies, a part of an ongoing project highlight seminaries in processing emotions in spoken languages across languages and cultures. Namely, both English speakers and Hebrew speakers are similarly biased to the emotional prosody of a spoken sentence, over its emotional semantics (Study 1).

The unique role of prosody across cultures, was also demonstrated, as German native speakers were able not only to correctly identify prosodic emotions in Hebrew, but also their misattributions of the prosodic cues paralleled those of native Hebrew speakers (Study 2). Finally, this prosodic dominance was found to occur in both a listener's native language (L1) and non-native language (L2, Study 3). Interestingly, it appears that in L2 prosodic dominance is even more pronounced. This effect may emanate from the apparent ease with which listeners process prosodic cues cross-linguistically (on auditory processing under cognitive stress, see Hadar, Skrzypek, Wingfield, & Ben-David, 2016). In sum, our data points on the universality and the special role prosody has in processing of emotions in spoken language.

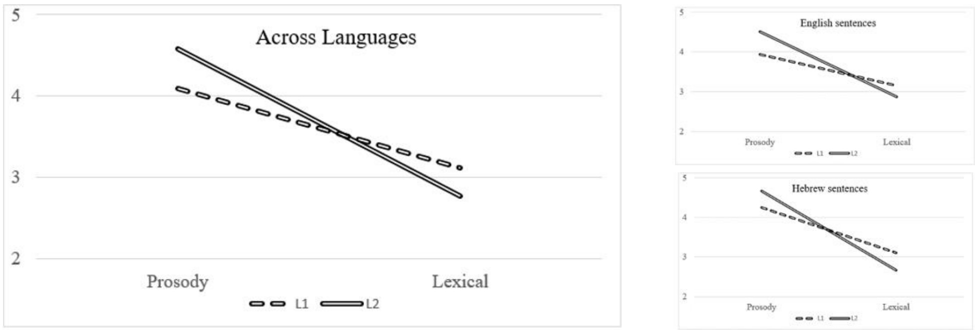


Figure 3. The extent of prosodic dominance (increased rating of prosody over semantic content in the general rating task) in a listener’s first language (dashed line, L1), and second language (solid line, L2). The main panel presents an average across both native Hebrew and native English speakers, the two side panels present the same data separately for each group.

Acknowledgements

This research was partially supported by a German Israeli Foundation grant (I-1324-105.4/2015) awarded to Ben-David Boaz M., Lachmann T., and Allen S. and a Marie Curie Career Integration Grant awarded to Ben-David B. by the European Commission (FP7-PEOPLE-2012-CIG).

References

Ben-David, B. M., & Algom, D. (2009). Species of redundancy in visual target detection. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(4), 958.

Ben-David, B. M., Avivi-Reich, M., & Schneider, B. A. (2016). Does the degree of linguistic experience (native versus nonnative) modulate the degree to which listeners can benefit from a delay between the onset of the maskers and the onset of the target speech? *Hearing Research*, *341*, 9–18.

Ben-David, B. M., Eidels, A., & Donkin, C. (2014). Effects of aging and distractors on detection of redundant visual targets and capacity: do older adults integrate visual targets differently than younger adults? *PloS One*, *9*(12), e113551.

Ben-David, B. M., Multani, N., Shakuf, V., Rudzicz, F., & van Lieshout, P. H. (2016). Prosody and semantics are separate but not separable channels in the perception of emotional speech: Test for rating of emotions in speech. *Journal of Speech, Language, and Hearing Research*, *59*(1), 72–89.

Ben-David, B. M., Thayapararajah, A., & van Lieshout, P. H. (2013). A resource of validated digital audio recordings to assess identification of emotion in spoken language after a brain injury. *Brain Injury*, *27*(2), 248–250.

Ben-David, B. M., van Lieshout, P. H., & Leszcz, T. (2011). A resource of validated affective and neutral sentences to assess identification of emotion in spoken language after a brain injury. *Brain Injury*, *25*(2), 206–220.

Hadar, B., Skrzypek, J. E., Wingfield, A., & Ben-David, B. M. (2016). Working memory load affects processing time in spoken word recognition: evidence from eye-movements. *Frontiers in Neuroscience*, *10*, 221.

Ishii, K., Reyes, J. A., & Kitayama, S. (2003). Spontaneous attention to word content versus emotional tone: Differences among three cultures. *Psychological Science*, *14*(1), 39–46.

Kitayama, S., & Ishii, K. (2002). Word and voice: Spontaneous attention to emotional utterances in two languages. *Cognition & Emotion*, *16*(1), 29–59.

Min, C. S., & Schirmer, A. (2011). Perceiving verbal and vocal emotions in a second language. *Cognition & Emotion*, *25*(8), 1376–1392.

Paulmann, S., & Uskul, A. K. (2014). Cross-cultural emotional prosody recognition: Evidence from Chinese and British listeners. *Cognition & Emotion*, *28*(2), 230–244.

Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics*, *37*(4), 417–435.

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, *32*(1), 76–92.

Shakuf V., Gal Rosenblum S., & Ben-David B. M,. (2016). The psychophysics of aging. In emotional speech, older adults attend to semantic, while younger adults to the prosody. In *Skotnikova, I., Korolkova, O., Blinnikova, I., Doubrovski, V., Shendyapin, V., & Volkova, N. (2016). Fechner Day 2016– Proceedings of the 32nd Annual Meeting of the International Society for Psychophysics.*

SESSION V

THE 2-D MIND: HOW CHILDREN DEVELOP AN UNDERSTANDING OF LINEAR PERSPECTIVE

Michelle Galanter
Teachers College, Columbia University

The development of pictorial depth understanding has not yet been completely explained. Though Hochberg and Brooks' groundbreaking study demonstrates that even babies can understand the meaning of pictures, we have not yet fully uncovered the process by which children come to know the relationship between the 3-dimensional world and 2-dimensional representations of it. Central to this issue is the anomaly of the perception of linear perspective. The ability to render linear perspective is both late to develop in childhood and was late to develop in human history. Until around age eleven, children are neither able to learn nor even able to retain information on how to draw using linear perspective. Likewise, until the Renaissance, the art world did not know how to render linear perspective. The perceptual implications of the difficulties apparent in linear perspective have remained unclear. To shed light on this question of the development of perceptual processing of linear perspective, children ages 8,9, and 10 were interviewed about stimulus pictures and then given one of two different interventions to accelerate their apprehension of linear perspective. The stimulus pictures were photographs of outdoor scenes involving buildings. The interventions were either to watch a video of a dolly-zoom effect on a cube or to draw a cube onto a piece of glass with a marker from a solid cube behind the glass. Children showed differential effects of the usefulness of the interventions on linear perspective understanding relative to age. Children also gave evidence of understanding linear perspective better when finger tracing was used to accentuate converging lines. The relevance of perceptual aids such as moving images, glass frames for drawing and gesture is noteworthy because of its implications for the teaching of spatial skills which are essential to a deep understanding of mathematics.

Reference

Hochberg, J. & Brooks, V. (1962). *Pictorial recognition as an unlearned ability: A study of one child's performance*. The American Journal of Psychology.

DIFFERENTIAL CONTEXTUAL LEARNING EFFECTS IN PHOTOPIC AND MESOPIC ENVIRONMENTS

Zhuanghua Shi[1], Xuelian Zang[1], Hermann J. Müller[1], and Linyun Huang[2]
[1] *Experimental Psychology, Department of Psychology, LMU Munich, Germany*
[2] *Department of Educational and Counselling Psychology, McGill University, Montreal, Canada*

Invariant spatial context can facilitate visual search, yet the question whether the learning and/or retrieval of context cues depends on luminance contrast and the ambient-lights is not well understood. Here, we conducted six visual search experiments under different combinations of luminance contrast and ambient-lights to investigate variations in contextual cueing. We found that context cues acquired with high-contrast displays did not transfer to low-contrast displays, irrespective of ambient-lights at learning and test (Experiments 1 and 3 with photopic and mesopic lighting respectively). Interestingly, learning of low-contrast contexts in a mesopic environment was possible (Experiments 4 and 6), and the learned contextual memory could be successfully used to guide search in high-contrast displays in the same, mesopic environment (Experiment 4), as well as in low-contrast displays in a photopic environment (Experiment 6). The transfer effect in the latter, albeit reduced, was still significant. By contrast, learning of low-contrast contexts in the normal, photopic environment turned out to be hard, in that no significant contextual-cueing effect emerged within five epochs of training (Experiment 2), and no transfer to high-contrast displays in the test phase (though some 'positive learners' did show a transfer effect). Taken together, these findings suggest that luminance contrast and ambient lighting interact with each other in contextual cueing, and mesopic (but not photopic) vision may boost contextual learning and retrieval of low-contrast contexts.

THE ROLE OF STEREOSCOPIC CUES IN IDENTIFYING OBJECTS IN A CLUTTERED VISUAL FIELD

Bruce A. Schneider and Cristina D. Rabaglia
University of Toronto Mississauga, Ontario, Canada
bruce.schneider@utoronto.ca; cris.rabaglia@gmail.com

We investigated the role of depth cues in identifying a letter presented at the point of fixation, when the target letter was embedded in a field of other letters perceived to be at the same distance from the observer, or appeared to be either in front of, or behind the surrounding letters. Surprisingly, the perceived position of the target letter relative to the depth plane of masking letters had no effect on performance. However, when the target letter was masked by visual noise having the same spectral content as the letter masker, target letters perceived to be in front of the noise plane were easier to identify than target letters behind the plane, which in turn were easier to identify than target letters on the same plane as the masking noise. The implications of these findings for object identification in a cluttered field will be discussed.

INFLUENCE OF EXPECTATIONS ON THE APPRECIATION OF BRIEF MOVIES

Joanie Drouin, Vincent Laflamme, and Simon Grondin
Université Laval, Québec, Canada

Abstract

This study is concerned with hedonic contrasts, that is, when a positive stimulus is judged more positively when a negative stimulus precedes it, and vice versa. This effect has been demonstrated for visual stimuli (images) and auditory stimuli (musical extracts). The present study uses audiovisual stimuli (4-minute animated short films) to verify the presence of hedonic contrasts. 80 students from Laval University participated in the experiment and were divided into 4 groups. They watched two short films, one classified as positive and the other as negative. After each viewing, they rated their level of appreciation using a score from 1 to 10. Participants in Groups 1 and 3 watched the negative film and then the positive film. The order was reversed for Groups 2 and 4. No specific instruction was given to participants of Groups 1 and 2. In contrast, participants of Groups 3 and 4 were given indications relative to the fact that the same film technique was or was not used for each film, and whether they were or were not produced by the same director. The results revealed that the appreciation scores of the groups which received no instructions (Groups 1 and 2) was significantly lower than those that did (Groups 3 and 4). However, this increase was significantly larger for the negative film than the positive film.

A beautiful painting is perceived even more positively when it is presented after a painting considered ugly, and vice versa (Zellner, Jones, Morino, Cogan & Jennings 2010). Such an effect of the order of presentation is called hedonic contrast. Hedonic contrasts have been demonstrated with images (Rota & Zellner, 2007; Zellner & al., 2009), photos (Cogan, Parker & Zellner, 2013; Hayn-Leichsenring, Kloth, Schweinberger & Redies 2013; Tousignant & Bodner, 2014), paintings (Zellner & al., 2010), food (Conner, Land, & Booth, 1987; Lahne & Zellner, 2015) and musical extracts (Parker, Bascom, Rabinovitz & Zellner, 2008).

Hedonic contrasts usually appear when the compared stimuli belong to the same category (Parducci & Perrett, 1971; Parducci, Knobel & Thomas, 1976). However, when stimuli are classified in separate categories, then no comparison is made, which prevents the creation of hedonic contrasts (Rota & Zellner, 2007; Zellner, Dawn & Henley, 2006). The categorisation effect is often observed when participants are considered ‘experts’ in a domain. For example, in the study of Rota and Zellner (2007), participants were asked to rate images of flowers (positive iris and negative orchids). They were divided into two groups: horticulture experts and non-experts. Results revealed that experts distinguished positive iris from less pleasant orchids, so they classified them in two different categories. As they did not compare flowers, no hedonic contrast was observed in their rating scores, contrary to the non-experts.

The present study is the first to verify the presence of hedonic contrasts with audiovisual stimuli (animated short films). Moreover, this study aims to document the effect of categorization on hedonic contrast.

Method

Eighty participants (36 women and 44 men, mean age = 25.11) were recruited from Université Laval and randomly assigned to one of four groups. All participants watched two 4-minute animation films, one rated as positive (Voloshin & Djumaev, 2012) and the other as negative (Firth, 2008) in a pilot project. After each viewing, participants rated their level of appreciation on a scale ranging from 1 to 10. Participants in Groups 1 and 3 watched the positive film first and participants in Groups 2 and 4 viewed the negative film first. To verify the presence of a hedonic contrast, no specific instruction was given to participants of Groups 1 and 2. The objective, with Groups 3 and 4, was to create a categorization effect and to see its influence on the manifestation of hedonic contrasts. So, the goal was to place participants in a position where they would be more likely to classify the movies in the same or two different categories. Participants of Group 3 were given indications that the same animation technique was used for both films and that they were produced by the same director. Participants of Group 4 were told that different animation techniques were used for the films and that they were produced by different directors.

Results

The mean and standard error of the appreciation scores for each movie and each experimental group are presented on Figure 1. A 2 (Movie: Positive vs Negative) x 2 (Order: Positive First vs Negative First) x 2 (Instructions: Without Descriptive Text vs With Descriptive Text) mixed design ANOVA was conducted on the appreciation scores, with Movie as a within-subjects factor and Order and Instructions as between-subjects factors. With a significance level of 5%, results showed a statistically significant main effect of Movie [$F(1, 75) = 82.90$, $p < .001$, $\eta^2_p = .525$], meaning that overall, participants attributed better scores for the Positive Movie ($M_{Score} = 7.47$, $SD = 1.78$) than for the Negative Movie ($M_{Score} = 4.72$, $SD = 2.48$). Moreover, the ANOVA revealed a significant main effect of Instructions [$F(1, 75) = 29.28$, $p < .001$, $\eta^2_p = .281$], which means that the participants whose instructions contained a descriptive text (Groups 1 and 2; $M_{Score} = 5.23$, $SD = 1.49$) gave overall lower appreciation scores than the participants whose instructions did not (Groups 3 and 4; $M_{Score} = 6.93$, $SD = 1.31$). These two findings were not independent from each other, however, since the ANOVA also revealed a significant Instructions x Movie interaction effect, $F(1, 75) = 6.54$, $p = .013$, $\eta^2_p = .080$. Follow-up tests on the simple effects show that the difference in appreciation scores between the Positive and Negative movies was greater for the groups whose instructions did not contain a descriptive text [Positive: $M_{Score} = 7.00$, $SD = 1.93$; Negative: $M_{Score} = 3.46$, $SD = 2.16$; $F(1, 37) = 59.59$, $p < .001$, $\eta^2_p = .617$] than for the groups whose instructions did [Positive: $M_{Score} = 7.92$, $SD = 1.51$; Negative: $M_{Score} = 5.94$, $SD = 2.14$; $F(1, 38) = 24.77$, $p < .001$, $\eta^2_p = .395$]. Moreover, the increase in appreciation scores brought by the inclusion of a descriptive text was larger for the Negative movie [Without Descriptive Text: $M_{Score} = 3.46$, $SD = 2.16$; With Descriptive Text: $M_{Score} = 5.94$, $SD = 2.14$; $F(1, 75) = 27.12$, $p < .001$, $\eta^2_p = .266$] than for the Positive movie [Without Descriptive Text: $M_{Score} = 7.00$, $SD = 1.93$; With Descriptive Text: $M_{Score} = 7.92$, $SD = 1.51$; $F(1, 75) = 5.51$, $p = .022$, $\eta^2_p = .068$]. All the other effects in the mixed design ANOVA failed to reach significance level (all ps > .15, η^2_p <.027).

Discussion

This study set on verifying the manifestation of hedonic contrasts and the categorization effect with animated short films. The results obtained suggest that the main effect of the order of presentation is not significant, which is rather surprising. Unlike auditory or visual stimuli (Parker & al., 2008, Zellner & al., 2010), no hedonic contrast was observed with audiovisual stimuli. Several features related to the short films used in this study could explain this result. First, they lasted about 4 minutes each, so it is possible to think that the temporal factor plays a crucial role in hedonic contrasts. In most of the previous studies, the visual stimuli presented to the participants had a duration of, at most, a few seconds (Forsythe, Zellner, Cogan, & Parker, 2014; Rota & Zellner 2007, Tousignant & Bodner 2014), or up to 20 seconds in the case of musical stimuli (Parker & al., 2008). Hedonic contrasts may fade out with the duration of the stimuli, which may make it more difficult to observe. Second, short films might be too complex to be influenced by the order of presentation. In fact, cinematographic works are characterized by several elements that can modulate appreciation (story, music, animation technique, sound design, etc.). Third, the absence of hedonic contrast in this study could be due to categorisation effect. Our results suggest that short films may be automatically categorized into different categories. Participants might already be experts because they have been exposed to a numerous amount of animation films in their youth. It is possible that this expertise my have allowed the participants to distinguish the different animation techniques (Stop Motion vs. computer animation) used in the two films, thus categorizing them separately and therefore not comparing them.

Although no hedonic contrast was found in this study, there was a significant interaction effect between the Instruction and Movie factors. This research suggests that simple descriptions attached to the films (Groups 3 and 4) are enough to modify the appreciation level. Positive or negative expectations might be built quickly after reading these descriptions, biasing favorably or unfavorably the short film to be seen (Zellner, Strickhouser, & Tornow, 2004). If short descriptions can modulate the appreciation, it would be relevant to verify how much ratings, used as a marketing strategy in the field of cinema for advertising movies, can influence the appreciation.

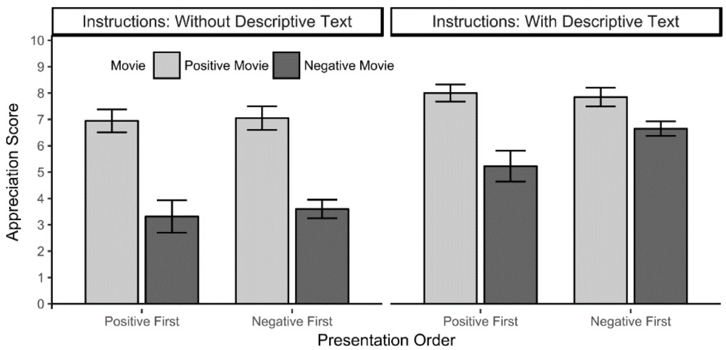


Figure 1. Mean and Standard Error for the Appreciation Scores as a function of the Presentation Order of the movies (Positive First for Groups 1 and 3 and Negative First for Groups 2 and 4) and of the presence (Groups 3 and 4) or absence (Groups 1 and 2) of a descriptive text in the instructions given to the participants. Error Bars correspond to 1 Standard Error of the mean.

References

Cogan, E., Parker, S. & Zellner, D.A. (2013). Beauty Beyond Compare: Effects of Context Extremity and Categorization on Hedonic Contrast. *Journal of Experimental Psychology: Human Perception and Performance, 39* (1), 16 – 22.

Conner, M.T., Land, D.G. & Booth, D.A. (1987). Effect of stimulus range on judgements of sweetness intensity in a lime drink. *British Journal of Psychology, 78*, 357-364.

Firth, D. (Producer and scenarist). (2008). *Crooked rot* [Short-films]. United Kingdom.

Forsythe, M., Zellner, D, A., Cogan, E. & Parker, S. (2014). Attractiveness difference magnitude affected by context, range, and categorization. *Perception*, *43*, 59-69.

Hayn-Leichsenring, G.U., Kloth, N., Schweinberger, S.R, & Redies, C. (2013). Adaptation effects to attractiveness of face photographs and art portraits are domain-specific. *i-Perception*, *4*, 303–316. Doi:org/10.1068/i0583

Lahne, J. & Zellner, D.A. (2015). The great is the enemy of the good : Hedonic contrast in coursed meal. *Food Quality and Preference, 45*, 70-74.

Parducci, A. & Perrett, L.F. (1971). Category rating scales: effects of relative spacing and frequency of stimulus values. *Journal of Experimental Psychology, 89*(2), 427-452

Parducci, A., Knobel, S. & Thomas, C. (1976). Independent contexts for category ratings: A range-frequency analysis. *Perception & Psychophysics, 20*(5), 360-366.

Parker, S., Bascom, J., Rabinovitz, B. & Zellner, D. (2008). Positive and negative hedonic contrast with musical stimuli. *Psychology of Aesthetics, Creativity, and the Arts, 2*, 171–174. doi:10.1037/1931-3896.2.3.171

Rota, M. & Zellner, D.A. (2007). The categorization effect in hedonic contrast: Experts

Tousignant, C. & Bodner, G.E. (2014). Context effects on beauty rating of photos: Building contrast effects that erode but cannot be knocked down. *Psychology of Aesthetics, Creativity, and the Arts*, *8* (1), 81-86.

Voloshin, D. (Producer) & Djumaev, S. (Scenarsit). (2012). *Dji. Death Fails* [Short-films]**.** Moldavie.

Zellner, D.A., Dawn, A. & Henley, M. (2006). Hedonic contrast and condensation: Good stimuli make mediocre stimuli less good and less different. *Psychonomic Bulletin & Review, 13* (2), 235-239.

Zellner, D.A., Jones, K., Morino, J., Cogan, E., Jennings, E. & Parker, S. (2010). Increased hedonic differences despite increases in hedonic range. *Attention, Perception, & Psychophysics, 72*(5), 1261-1265. doi : 10.3758/APP.72.5.1261

Zellner, D.A., Mattingly, M.C. & Parker, S. (2009). Categorization reduces effect of context on hedonic preference. *Attention, Perception, & Psychophysics, 71*(6), 1228-1232. doi:10.3758/APP.71.6.1228

Zellner, D.A., Strickhouser, D. & Tornow, C.E. (2004). Disconfirmed hedonic expectations produce perceptual contrast, not assimilation. *American journal of psychology, 117* (3), 363-387.

BOUNDARY EXTENSION AND THE CUEING OF SCENE CATEGORY

Timothy L. Hubbard
Department of Psychology, Arizona State University, Tempe, AZ USA and College of Doctoral Studies, Grand Canyon University, Phoenix, AZ USA
Susan E. Ruppel
Department of Psychology, University of South Carolina Upstate, Spartanburg, SC USA
timothyleehubbard@gmail.com

Whether boundary extension in memory for a scene could be influenced by prior expectations regarding the semantic category of that scene was examined. In all experiments, observers viewed a scene containing either a bird, car, church exterior, kitchen interior, or natural landscape. In Experiment 1, 50% of participants received a one-word verbal (written) cue identifying the category of the upcoming scene, and the other 50% of participants did not receive a cue. In Experiment 2, all participants received a cue identifying the category of the upcoming scene, and this cue was valid on 80% of the trials and invalid on 20% of the trials. In Experiment 3, no cues were presented; for 50% of participants, scenes were blocked by category, and for the other 50% of participants, scenes were ordered randomly. Boundary extension occurred in each condition in each experiment, but was not influenced by cueing or by blocking of scene category. Also, the magnitude of boundary extension was influenced by scene category. It is suggested that semantic information regarding a scene category is neither necessary nor sufficient for boundary extension, but that such information might modulate boundary extension. Implications for theories of boundary extension and for relationships of boundary extension to attention, semantic knowledge, and scene priming are considered.

THE PSYCHOPHYSICAL BIOMARKER OF CHRONIC STRESS: THE DESYNCHRONIZATION OF MAGNOCELLULAR AND PARVOCELLULAR VISUAL PATHWAYS

Irina Shoshina[1], Roman Sergienko[2], Ekaterina Zavyalova[2], and Andrey Gruzdev[2]
[1] *Pavlov Institute of Physiology Russian Academy of Sciences, St. Petersburg, Russia;*
[2] *Siberian Federal University, Krasnoyarsk, Russia*
shoshinaii@mail.ru

Introduction

The aim of this study was to examine the functional state of magnocellular and parvocellular visual pathways, features of their interaction in a chronic stress condition on model of professional burnout. The magnocellular system is more sensitive to the low spatial and high temporal frequencies, and they promote rapid information transmission to the neurons, mostly of the dorsal pathway. Because of these properties, the magnocellular neurons are involved in processing of information on the global organization of a stimulus and in analysis of movement of the observed object. The parvocellular system is more sensitive to the high spatial and low temporal frequencies. They provide a slower information transmission to the neurons of primarily ventral pathway. The parvocellular system is responsible for description of color and fine details of the object. The interaction of magnocellular and parvocellular visual systems provides the integrity of perception and adaptive behavior.

Methods

Twenty one university employees and twenty university students participated in the study. Burnout was measured with diagnostic methods of the emotional burnout (V.V. Boyko). Contrast sensitivity thresholds was measured with the visual contrastometry (Gabor elements with spatial frequencies of 0.4, 3.6 and 17.8 cycles/degree were presented).

Results

Between-group comparisons revealed a statistically significant difference between groups (participants without burnout, participants with the resistance phase at a formative stage and participants with symptoms of formed resistance phase) with respect to contrast sensitivity thresholds at low ($F(2,205)= 3.03$; $p = 0.05$), medium ($F(2,238)= 11.67$; $p < 0.001$) and high spatial frequencies ($F(2,200)= 5.11$; $p = 0.05$). A post-hoc Games-Howell test revealed that participants with the resistance phase at a formative stage demonstrated increased contrast sensitivity at low spatial frequencies compared to individuals without burnout. Compared to individuals without burnout, participants with symptoms of formed resistance phase had reduced contrast sensitivity at medium and high spatial frequencies. Also, participants with symptoms of formed resistance phase showed reduced contrast sensitivity at medium and high spatial frequencies compared to participants with the resistance phase at a formative stage.

Conclusions

Thus, we have demonstrated the important role of matched functioning of magnocellular and parvocellular visual pathways in adaptive behavior in burnout. We propose to consider the functional state of magno- and parvosystems as a biomarker of chronic stress.

Additional Information

This research was supported by Krasnoyarsk State Autonomous Institution «Krasnoyarsk Regional Foundation of Research and Technical Activities» as part of the internship in 2017: «The methods for assessment of the functional state of visual perception mechanisms in depression and burnout» (Grant No. 2017052402493).

SESSION VI

THE LOCUS OF DECISION-MAKING AND CONFIDENCE PROCESSING: AUTONOMOUS OR INTEGRATED RESPONSE SELECTION SYSTEMS?

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

Abstract

The present study examines three features of uncertainty monitoring in perceptual judgement: the number of mental processes involved, the locus of confidence processing, and the sources of evidence used to assess certainty. Participants performed a line bisection task and provided confidence judgments before (prospective) or after (retrospective) primary decision response selection. Response time analyses indicated that confidence processing reflected an additive model based on the primary decision and confidence judgment. In contrast, calibration analyses were not affected by confidence response order but were affected by flanker congruency and task difficulty. Consequently, the present study suggests that performance monitoring reflects a secondary set of processes that are affected by sources that are distinct from those used to perform the primary decision.

Uncertainty monitoring appears to be a feature of human and nonhuman cognition (Smith et al., 2017). Confidence reports are often used in perceptual identification, discrimination, and categorization tasks as a means to assess participants' awareness of their performance. Despite the widespread use of confidence reports, there are few formal theoretical frameworks that have attempted to integrate evidence for the relationship between the processes and representations that participants use to monitor their performance (e.g., Baranski and Petrusic, 1998; Busey, et al., 2000; Keren, 1991; Nelson & Narens, 1990). The current study assess the number of processes, the locus of confidence, and the evidence that affect certainty in order to provide a coherent framework to understand confidence.

Metacognition in Perceptual Judgments: Processing and Sources of Evidence

Number of Processes.

Confidence reports have been be considered in terms of whether they are a byproduct of the primary decision (*direct-scaling*) or whether they require additional processing (*indirect-scaling*). With the introduction of signal-detection theory (SDT), confidence processing was understood as an effortless byproduct of the primary decision (Ferrel & McGooey, 1980). For instance, in a two-alternative forced-choice task, a criterion is used to partition perceptual space. The location of a stimulus signal relative to this criterion would determine primary decision response selection. Confidence reports would be generated in a similar manner by adding additional criterion: as the relative distance of the stimulus from the decision criterion increased (i.e., the strength), the level of confidence also increases. Consequently, no additional processing events are required in order to produce confidence judgments. Early models of metamemory also shared similar assumptions (Hart, 1965): the strength of a memory trace would determine both the accuracy of recall as well as the participants confidence. Subsequent models assumed confidence reports were obtained indirect from the primary decision through additional processes (e.g., Audley, 1960; Baranski & Petrusic, 1998; Vickers, 1979). These processes introduce errors and systematic biases as a result of the scaling operations leading to miscalibration and under- or overconfidence.

Locus of Confidence

With the introduction of indirect-scaling models, confidence reports could either be processed during (*decisional-locus*) or after the primary decision (*postdecisional-locus*). SDT-based models of confidence processing necessarily assumed a that confidence processing was an extension of the primary decision. Subseqeunt SDT-based models have used this same principle to account for difference is confidence reponse time (CRT): confidence reflects a second decision generated from the same decision-making process (cf. Pleskac & Buseymeyer, 2010). Heuristic models of decision-making have also suggested a decision locus, such that a cue is used to make a decision while the estimated cue validity determines subjective confidence (Gigerenzer et al. 1991). In contrast to these models, most indirect-scaling models assume that rescaling of primary decision evidence occurs post-decisionally (Vickers, 1979) or that confidence has an alterable locus (Baranski & Petrusic, 1998). For instance, Baranski and Petrusic (1998) found that participants' CRT in a speed-stress condition was longer than their CRT in an accuracy-stress condition. With the opposite pattern evidenced for decision response time, they interpreted these results as evidence of an alterable locus that was dependent on primary decision task demands.

Sources of Evidence

Indirect-scaling models also introduce the possibility that confidence reports could simply be based on evidence used during the primary decision (*target-based*) or additional cues or non-diagnostic information (*cue-based*). Target-based models of confidence processes are well represented in early confidence models. Participants' confidence judgment was believed to be the result of either the signal strength (Ferrel & McGooey, 1980), the number of vacillations between dominant and nondominant responses (Audley, 1960), the amount of accumulated evidence for the dominant response alternative (Vickers, 1979). Baranski and Petrusic (1998) additionally suggested that participants might examine the quantity of nondiagnostic information they had accumulated over the course of the trial rather than the diagnostic evidence that favoured one of the response alternatives. In their model, accumulated nondiagnostic information reflects a participant's doubt in their decision. Metacognitive models of general knowledge (Gigerenzer et al, 1991) and metamemory (Koriat, 1993) have instead suggested that nondiagnostic information in the form of stimulus cues can influence judgments. Notably, Koriat and colleagues (e.g., Koriat et al., 2002) found that metamemory judgments were influenced by the ease of encoding or retrieval. Encoding cues would be used if metacognitive judgments were made shortly after memory items were studied whereas retrieval cues would be used when there were lengthier delays between studying and judging their learning performance. Thus, relative to the primary decision both target-based information and cue-based information appear to influence subjective assessments of performance (e.g., Schoenherr, Leth-Steensen, & Petrusic, 2010).

Present Study

The current study attempts to examine whether confidence reports are scaled indirectly from diagnostic and nondiagnostic information available to participants and whether confidence processes reflect a separable set of mental operations. Using a line bi-section task, the current study manipulated the presence and location landmarks (X or Xs) that flank a target line. Relative to when no landmarks or bilateral landmarks are used, the use of a unilateral landmark introduce perceptual bias into perceptual judgments when participants must decide which side is longer/shorter. Consequently, this might differentially affect the primary decision and confidence response processes.

In order to assess when confidence processing occurred, I additionally manipulated whether participants were required to report confidence prospectively or retrospectively. When making a retrospective confidence report, participants should have already completed the primary decision and should therefore only need to rescale the available evidence to report their confidence. However, making a prospective confidence report, participants would still need to first make a decision in order to report confidence if confidence reports are dependent on the primary decision then a significant increase in confidence response time should be observed. Using additive factors logic (Sternberg, 1998), this hypothesis can be explored directly. Namely, prospective confidence should be predicted by:

$$\text{Additive Model } (CRT_P) = DRT_{NC} + CRT_R$$

Where prospective CRT, CRT_P, is estimated by obtaining the mean of the primary decision in the no confidence condition, DRT_{NC}, and the mean of confidence response time in the retrospective confidence condition, CRT_R. Thus, the additive model suggests that the primary decision and confidence process are separate but dependent processes. Finally, given that prospective confidence reports require the production of a decision before confidence can be provided, I assume that confidence calibration indices should be relatively unaffected in comparison to the retrospective confidence condition.

Method

The stimuli consisted of bisected lines 1-pixel in height. Three stimulus difficulty conditions created by manipulating the difference between the longer and shorter size by 1 (hard), 2 (intermediate), and 3 (easy) pixels. Four flanker conditions were created by including no flankers, one flanker on the left or right of the stimulus, or on both sides of the stimulus. When present, an ”X” was used as a flanker and was located 1 cm away from bisected line. Participants were required to select which side was larger or smaller. Confidence reports were obtained using a 6-point response scale with range 50 (guess) to 100 (certain). Confidence response were either required before primary decision response selection (prospective) or following the completion of the primary decision (retrospective).

Results and Discussion

While asymmetric effects are observed in unilateral flankers in line bisection tasks, left and right unilateral flanker conditions were collapsed, creating three landmark conditions: no flankers, a unilateral flanker, and bilateral flankers. Thus, subsequent repeated measures ANOVA consisted of a 3 (landmark location) x 3 (confidence order) x 3 (stimulus difficulty) design. Due to a coding error, other response keys on the keyboard were not suppressed leading participants to press these keys erroneously. This accounted for only 1.2% of responses. These trials were removed prior to the analyses.

Proportion Correct

The ANOVA analysis of proportion correct revealed significant differences for the effect of stimulus difficulty, $F(2, 56) = 155.782$, $MSE = .015$, $p < .001$. A monotonic decrease in performance was associated with increases in discrimination difficulty. Bonferroni *post-hoc* analyses revealed that all difficulty conditions differed from each other (all *p*s < .001). The effect of landmark location was also significant, $F(2, 56) = 6.329$, $MSE = .024$, $p = .004$. Bonferroni *post-hoc* analyses revealed that this difference was a consequence of the no flanker differing from the unilateral flanker condition ($p = .002$). Importantly, there was no effect of confidence order on accuracy, $F(2, 56)=2.19$, $p = .127$. This means that neither prospective or retrospective confidence reports facilitated or interfered with the performance in the discrimination task.

Decision Response Time

Prior to analysis, DRTs that were 3 SD greater than the mean were removed from the analyses. The ANOVA for DRT obtained a main effect of confidence order, $F(2,56) = 36.195$, $MSE = 377675$, $p < .001$. Bonferroni *post-hoc* analyses revealed a significant difference (all *p*s < .001) between DRT in the prospective confidence condition and both the no confidence and the retrospective confidence conditions (see Table 1). A main effect of discrimination difficulty was also obtained, $F(2, 56) = 15.057$, $MSE = 53434$, $p < .001$. Bonferroni *post-hoc* analyses revealed a significant (all *p*s < .004) difference between the easy condition (1086 ms) and intermediate (1154 ms), and difficult condition (1187 ms). Finally, a main effect of landmark location was also obtained, $F(2, 56) = 7.592$, $MSE = 62794$, $p = .003$. Bonferroni *post-hoc* analyses revealed a significant (all *p*s < .01) difference between the unilateral flanker condition (1186 ms) and the no flanker condition (1122 ms) and the bilateral flanker condition (1119 ms).

Table 1. Observed Response Time and Additive Model of Prospective Confidence Reports.

	No Confidence	Retrospective Confidence	Prospective Confidence	Additive Model
DRT	1004 ms	1414 ms	1041 ms	
CRT		641 ms	1587 ms	1645 ms

Confidence Response Time

Prior to analysis, CRTs that were 3 SD greater than the mean were removed from the analyses. The ANOVA of CRT obtained a main effect of confidence order, $F(1, 28) = 168.809$, $MSE = 665076$, $p < .001$. Confidence order also interacted with both discrimination difficulty, $F(2, 56) = 9.535$, $MSE = 42727$, $p < .001$, and landmark location, $F(2, 56) = 4.558$, $MSE = 47676$, $p = .019$. As Figure 1 demonstrates, while CRT was uniformly greater in the prospective confidence report condition, CRTs were also faster in the unilateral flanker condition. Thus, confidence processing time was affected by stimulus features that should already have been processed during the primary decision.

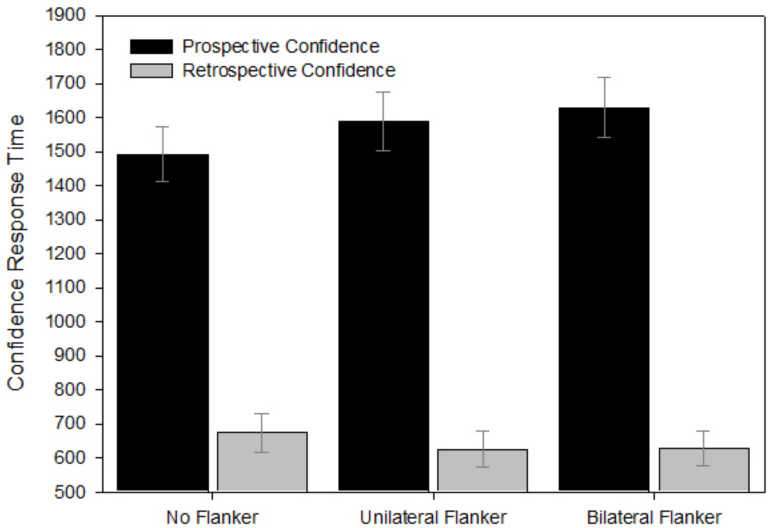


Figure 1. Interaction between landmark location and confidence order.

DRT and CRT Additive Model

As the method of the present study was a repeated-measures design, I can directly compare the additive model for each participant to their obtained results. While there was only a small difference in mean RTs for the additive model and obtained CRT_P ($M_{DIFF} = 58.36$ ms). As Table 1 demonstrates, the Additive Model of CRT_P closely approximated obtained CRT_P for each participant. A strong, positive correlation between the model and prospective confidence reports was also obtained, $r(29) = .564$, $p = .001$. Moreover, a paired-samples t-test did not reveal a difference between the additive model for CRT_P and the obtained CRT_P, $t(28) = -.733$, $p = .470$.

Thus, adequacy of the Additive Model suggests that confidence processing *requires* the output of the primary decision in order to scale confidence in perceptual judgments. In this information is not available, the participant must obtain it prior to reporting their subjective confidence.

Confidence Calibration Indices

Mean Confidence. The results of the ANOVA of mean confidence differed from that of the analysis of proportion correct. Mean confidence was significantly affected by the confidence response order, $F(1, 28) = 4.741$, $p = .038$. When participants were required to provide a prospective confidence report they were somewhat less confident ($M = 74.6$) than when they there were required to provide a retrospective confidence report ($M = 76.8$). A marginally significant effect was also observed between difficulty condition and landmark location, $F(4,112) = 2.648$, $MSE = 13.36$, $p = .054$. This qualifies the main effect of stimulus difficulty that was also significant, $F(2, 56) = 48.253$, $MSE = 39.742$, $p < .001$. Whereas in the easy condition confidence was unaffected by the landmark location, greater confidence was reported in the unilateral flanker conditions relative to the no flanker and bilateral flanker conditions. Importantly, this might reflect a Hard-Easy Effect as accuracy was the lowest in these conditions a possibility that I will examine below.

Confidence Calibration. Confidence calibration reflects the unsigned difference between a participant's accuracy and their confidence response on a trial-to-trial basis (Baranski & Petrusic, 1994). Larger values represent greater deviation from perfect calibration (e.g., a participant should say they are 70% when the proportion correct is .70). The ANOVA of confidence calibration revealed that the only significant factor to affect calibration was task difficulty, $F(2, 56) = 4.189$, $MSE = .001$, $p = .036$. Miscalibration increased as a monotonic function of task difficulty across the easy ($M = .067$), intermediate ($M = .073$), and hard difficulty conditions ($M = .088$). A marginal effect of confidence order was also observed, $F(1, 56) = 3.286$, $MSE = .003$, $p = .081$. Participants were marginally better calibrated in the retrospective confidence report condition ($M = .071$) than the prospective confidence report condition ($M = 0.81$). Thus, despite the need to make a primary decision prior to reporting confidence, prospective confidence reports might alter either the primary decision process or use different cues in order to report confidence.

Overconfidence. Overconfidence was examined by obtaining the difference between mean confidence and mean proportion correct. The resulting difference score indicate whether a participant was overconfident or underconfident relative to accuracy with a positive and negative value, respectively. The ANOVA for overconfidence revealed a main effect of task difficulty, $F(2, 56) = 60.388$, $MSE = .014$, $p < .001$. Participants expressed overconfidence in the hard condition ($M = .043$) and underconfidence in both the intermediate ($M = -.047$) and easy difficulty conditions ($M = -.083$). This pattern reflects the Hard-Easy Effect (Lichtenstein & Fischoff, 1977). Landmark location also affected overconfidence bias, $F(2, 56) = 8.202$, $MSE = .014$, $p = .001$. Bonferroni *post-hoc* comparisons revealed that the unilateral flanker ($M = -.002$) condition was associated with no underconfidence bias in comparison to the no landmark condition ($M = -.046$; $p = .003$) or the bilateral landmark condition ($M = -.039$; $p = .02$).

Conclusions

The present study appears to offer support for indirect scaling models of confidence processing with an alterable locus. First, increases in mean DRT in the retrospective confidence condition relative to the no confidence condition suggest that participants are likely assessing their performance *while* making a decision. Thus, the requirement of confidence reports induces concurrent performance monitoring while making a decision. Further, given the equivalence of mean DRTs for the no confidence condition and the prospective confidence condition, this suggests that decision processes that followed confidence reports were not being monitored. Confidence processing appears to reflect a secondary set of operations.

Second, the increase in CRT in the prospective confidence condition relative to the retrospective confidence condition suggests that participants were required to complete a primary decision *prior* to assessing the decision process or the evidence used therein. An additive factors model of confidence reports appears to account for these findings. Prospective confidence reports were not significantly different from the sum of the decision response time in the no confidence condition (DRT_{NC}) and confidence response time in the retrospective condition (CRT_R). Given that both of these conditions allow the primary decision and confidence processing to complete without additional interference, prospective confidence reports appear to require rescaling of output of the primary decision processes.

Finally, the present study provides evidence for the Hard-Easy Effect (Lichtenstein & Fischoff, 1977). Participants reported overconfidence in difficult conditions and underconfidence in easier conditions. I additional observed less underconfidence bias in the unilateral flanker condition relative to the no flanker and bilateral flanker conditions. Thus, unilateral flankers appear to increase confidence to counter the underconfidence observed in the other conditions. When considered with other evidence (Schoenherr et al., 2010), this suggest that while confidence reports are influenced by both target- and cue-based information.

Acknowledgements

Special thanks to Craig Leth-Steensen for the use of his laboratory space for the conducted of this experiment.

References

Audley, R. J. (1960). A stochastic model for individual choice behaviour. *Psychological Review*, 67, 1-15.

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.

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.

Ferrel, W. R., & McGoey, P. J. (1980). A model of calibration for subjective probabilities. *Organizational Behaviour and Human Performance*, *26*, 32-53.
Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, 98, 506-528.
Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology, 56*, 208–216.
Keren, G. (1991). Calibration and probability judgments: Conceptual and methodological issues. *Acta Psychologica, 77,* 217–273.
Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, 100, 609-639.
Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experiment Psychology: General*, 131, 147–162.
Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? The calibration of probability judgments. *Organizational Behaviour and Human Performance*, 20, 159-183.
Nelson, T. O., & Narens L. (1990). Metamemory: A theoretical framework and new findings. In G H Bower (Ed) *The Psychology of Learning and Motivation* (Vol 26 pp 125-173) New York Academic Press.
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., Leth-Steensen, C., & Petrusic, W. M. (2010). Selective attention and subjective confidence calibration. *Attention, Perception & Psychophysics*, 72, 353-368.
Smith, J. D. & Church, B. A. (2017). Dissociable learning processes in comparative psychology. *Psychonomics Bulletin & Review*, 1-20.
Vickers, D. (1979). *Decision processes in visual perception*. New York: Academic Press.

SAME STIMULUS – DIFFERENT CONSTELLATION: THE VISUAL MISMATCH NEGATIVITY INVESTIGATED WITH GARNER'S PATTERNS.

Ann-Kathrin Beck[1], Thomas Lachmann[1], Daniela Czernochowski[1], Corinna A. Christmann[1], and Stefan Berti[2]
[1] *University of Kaiserslautern, Kaiserslautern, Germany*
[2] *Johannes Gutenberg University of Mainz, Mainz, Germany*
Ann-Kathrin.Beck@sowi.uni-kl.de

In the present study we used the vMMN in order to investigate whether our mind differentiates stimuli based on perceptual differences even with unchanging sensory information. The visual Mismatch Negativity (vMMN), an event related potential (ERP) component, can be elicited by an infrequent deviant in a sequence of more frequent standards.

In rating studies, Garner and Clement (1963) used a set of dot patterns. Each consisting of five dots distributed on an imaginary 3x3 matrix leaving no row or column empty, consequently creating dot patterns with identical sensory energies and same physical entities. However, even if one pattern matches another after being rotated or mirrored, patterns are perceptually different.

For this instance, we recorded electroencephalographical (EEG) data in two studies with each 21 participants (average 25.9 and 25.6 years old). The dot patterns were presented in a 3-stimulus oddball-paradigm with an additional control, equiprobable condition. Participants were asked to count predefined target patterns in order to shift their attention away from the other patterns.

We compared one particular dot pattern across different conditions as defined by relative stimulus frequency and task instructions (serving as control condition, deviant or target stimulus). The vMMN was elicited in the comparisons between the control and deviant condition. Thus, our results suggest that the system underlying the vMMN can differentiate between perceptual modifications of stimuli, even if their sensory information remains constant. Additionally, we discovered no difference between the deviant and target condition. Hence, our results indicate that the underlying system of the vMMN is equally effective in differentiating stimuli with or without attention.

COMPARING SENSORY CAPACITY:
AN ESTIMATE OF FUNCTIONAL WEALTH

Christian Graff
Laboratoire de Psychologie et de Neuro Cognition, Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, 38000 Grenoble, France
Christian.Graff@univ-grenoble-alpes.fr

Defining a normalized score of sensory functional wealth may be convenient to quantitatively compare sensory systems across species (e.g. human vs animals) or modalities (e.g. vision vs hearing). The amount of stimuli that might be evoked from a physical continuum depends on the dynamic range on one hand, and on discrimination ability or precision on the other hand. In order to combine range and discrimination, I suggest expressing both of them as *geometric differences* on the physical scale of the continuum measurement. The geometric difference expresses the interval between two values by the natural logarithm of their ratio. Here, the dynamic range is the interval separating the two absolute thresholds; the precision is the reciprocal of the interval between two just differentiated stimuli, which is actually the reciprocal of the relative differential threshold. In its primary expression, the functional sensory wealth on a continuum scale of stimuli S is therefore approximated by the ratio of the dynamic range over the (mean) Weber fraction W, both expressed as geometric differences:

$$FSW = ln(S_{max}/S_{min})/W = ln(S_{max}/S_{min})/ln[(S_{ref}+jnd)/S_{ref}] \text{ (Fig.1)}$$

with *jnd* the just noticeable difference between a reference stimulus S_{ref} and a neighboring stimulus differentiated from it half the time, S and *jnd* both expressed in a SI physical unit.

For instance, we tested the electroreception ability of mormyrid fishes on two continua, that of resistive and capacitive properties of objects. To make sense of the results, we compared the two obtained electro-receptive functional wealth, to the human functional wealth of perception for luminance and for color, and for loudness and pitch. In line with the same principle, we attempted a similar quantitative comparison between human vision and hearing.

However, the Weber ratio W is not strictly constant on a given continuum. Accordingly, the above computation may be refined after a curve that presents precision (as the reciprocal of the Weber ratio) along the log-scale of the continuum S. The displayed sensory wealth is the area below the curve between the two absolute thresholds, i.e. the *integral of 1/W(S) dLn(S)*.

This computation of sensory wealth opens some avenues. 1) It may apply to effector systems as well as to sensory ones, and to artificial as well as to natural systems. 2) Multidimensional sensory *palettes* can be defined by combining several continua in a given system, such as loudness with pitch or luminance with color. However, physiological restrictions reduce the comprehensive use of the full theoretical product of two continua wealth to obtain the palette wealth of a natural system, whereas such product may directly apply to artificial sensors or actuators. 3) Finally, I consider converting or translating functional wealth of continua or palettes into units of information such as bits.

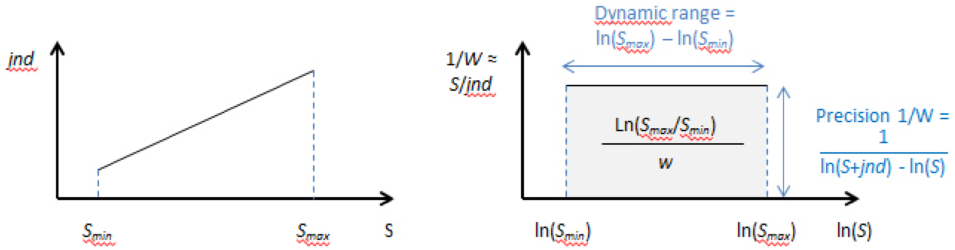


Fig. 1. Sensory wealth in grey for theoretical Weber's law, i.e. for $W = \text{Ln}(S+jnd)/S) \approx jnd/S$ constant.

REPTILES' DECISION MAKING STRATEGIES IN A CHOICE BETWEEN RESPONSE ALTERNATIVES (COLOR DISCRIMINATION TASK) AND POSSIBLE SIMILARITIES WITH HUMAN BEHAVIOR

Irina G. Skotnikova
Institute of Psychology Russian Academy of Sciences, Moscow, Russia
Roman V. Zhelankin
Moscow Institute of Psychoanalysis, Moscow, Russia
iris236@yandex.ru, zhelankin86@mail.ru.

Abstract

3 main parameters of decision making were investigated regarding animals' behavior (for reptiles firstly): accuracy, response time and reacting strategies. Grass snakes' (Natrix natrix') choice of 1 of 2 ways in a T-maze was studied in a task of colors (red — green) discrimination which was poorly known in reptiles. It was found that average erroneous responses time was greater than correct ones time in difficult conditions while vice versa in easy ones. It agrees with R. Swensson's (1972) rule established in humans. 4 individual kinds of reacting strategies revealed, were similar to impulsive, reflective, fast/accurate and slow/inaccurate ones in humans. Snakes checked that way in the maze more often, which they intended to choose. It corresponds to models of human decision making, pointed that a certain response alternative is chosen if the greatest evidence have been accumulated in favor of it. Thus, we suggest a number of animals' turns towards each way in a maze to be an indirect index of an evidence value accumulated in favor for this alternative.

Studies of decision making and their individual differences in humans and animals

Decision making (DM) is a key component of every human and animals activity act, in sensory tasks in particular. As a rule, DM in humans is understood to be a choice of one of response alternatives compared (Anokhin, 1974; Sokolov, 2003; Tversky et al. 2005). Preparation of such a choice may be rather long in a course of human activity and is considered as a pre-decision included 3 phases (Kozeletsky, 1979).

DM main parameters are: accuracy, latent time and a subject's hesitations (confidence — confidence in its process and a result). Erroneous responses are given more quickly than correct ones in easy sensory discrimination and/or an attitude for a fast response, while more slowly in difficult discrimination and/or an attitude for a correct response (Swensson, 1972). DM individual peculiarities are manifested in reflective/impulsive cognitive style which is estimated by J. Kagan's Matching Familiar Figures Test in which a single picture has been found among 8 similar ones, that is quite identical to a standard picture (Kagan, 1966). 4 patterns of behavior have been revealed according to 2 indices calculated on the base of data obtained: an average (for 6 experimental cards presentations) response time (RT) of the 1st response and a total (for the 6 cards) number of errors (EN). Namely: reflective pattern (RT is great and EN is few), impulsive (vice versa), fast/accurate (RT is short and EN is few) and slow/inaccurate (both RT and EN are great).

These characteristics describe human DM. At the same time a generality of human's and animals' activity act structure was established in the Theory of a Functional System (Anokhin, 1974). DM occupies a central place in this structure. According to this tradition, we use the Decision Making notion in order to describe an animal's choice between behavior alternatives. A correspondence between DM psychological regularities in humans and animals was revealed experimentally as well (Smith, Ratcliff, 2004).

Time of reward waiting was used as an indicator of confidence in rats who made an odor choice (Lak et al., 2014). Monkeys rejected to choose a movement direction in a noisy display if they were not sure in this choice correctness (Fetsch et al., 2014). Impulsiveness is studied to be one of the base individual features in humans and animals. Cognitive impulsiveness went along with low memory characteristics and poor switching over to another action in rats (like as in humans), while behavior impulsiveness vice versa (Zaichenko et al., 2016). Hence these rats might be fast/accurate but not impulsive — but they were not checked for 4 Kagan's patterns of behavior.

The works mentioned were made in mammals. Individual differences in reptiles (who were important evolutionary predecessors of birds and mammals) were studied in extrapolation tasks (Ochinskaya, 1990). Snakes and skinks took their bodies borders in account when choose an opening in a screen for a food achievement (Khvatov et al., 2015).

Methods of the experimental study.

We investigated firstly the characteristics mentioned above, simultaneously in grass snakes (Natrix natrix), who made colors discrimination task. Namely: 3 basic DM parameters (accuracy, latent time and reptile's pre-decisional movements, probably reflected it's external information perception, orientation and hesitations in the choice) together with 4 Kagan's behavior patterns analysis. Experiments were carried out in a T-maze consisted of the following components: a starting chamber (SC); a small corridor (SC); a large corridor (LC); two color cabins: green (GC) and red (RC) (Fig. 1). For the both cabins conditions were equalized regarding illumination, temperature, tactile characteristics, visual environment; an influence of an odor was eliminated.

Firstly, snakes were trained to choose a GC (but not RC), receiving a support by an unconditional stimulus (warmth of the floor (Safarov, 1990) of the GC, or a smell of food, or feeding there). An animal could not feel the warmth of the cabin when being in the middle of the LC in 30 cm from the cabins, since the grass snakes do not have distant thermo receptors. Snakes were launched into the starting chamber, an entrance into the maze was opened and their behavior was recorded by a video camera and a web camera. The experiment was completed after the snake's stopping in the selected cabin for more than 1 minute.

Each of 6 snakes participated in 14 experiments under ordinal conditions and in 9 ones under some difficulties. In each experiment the following characteristics were recorded.

1) 8 moving time values: intervals (in seconds) between a snake's entering into the LC and it's reaching of the correct cabin or the erroneous one, and the total travel time from the SC to the cabins: **1.1)** under **ordinal** conditions: in the LC: MT1(RC) — mean time to the RC, MT1(GC) — mean time to the GC; in summary: MT2(RC) — mean time to the RC, MT2(GC) — mean time to the GC; **1.2)** under **difficult** conditions: in the LC: MT3(RC) — mean time to the RC, MT3(GC) — mean time to the GC; in summary: MT4(RC) — mean time to the RC, MT4(GC) — mean time to the GC.

2) According the video registration, 8 values of a number of a snake's head and/or body turns were counted: towards the correct cabin and towards the erroneous one before the correct choice and before the erroneous one. **2.1)** Under **ordinal** conditions: in **correct** choices: MNT1(CC) — mean number of turns towards a cabin chosen, MNT2(CR) — mean number of turns towards a cabin rejected; in **erroneous** choices: MNT3(CC) — mean number of turns towards a cabin chosen, MNT4(CR) — mean number of turns towards a cabin rejected. **2.2)** Under **difficult** conditions: in **correct** choices: MNT5(CC) — number of turns towards a cabin chosen, MNT6(CR) — number of turns towards a cabin rejected; in **erroneous** choices: MNT7(CC) — number of turns towards a cabin chosen, MNT8(CR) — number of turns towards a cabin rejected.

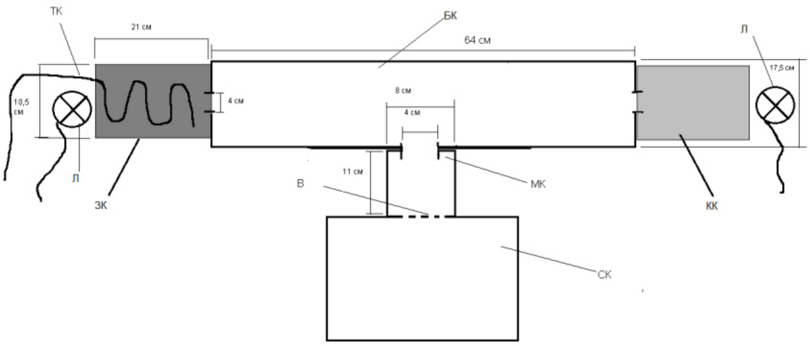


Fig.1. T-maze (the components marks see above).

A significance of differences between the indices calculated were assessed by Wilcoxon test. Individual data classification was made in order to check Kagan's behavior patterns regarding medians of erroneous choices proportions (EP) and time of snakes' staying in the LC (probably, it is time of a choice — TC). Individual data classification was made regarding medians of erroneous choices proportions (EP) and time of snakes' staying in the LC (probably, it is time of a choice TC).

Results

I. Time parameters

To test the Swensson's rule, data were compared that were obtained under ordinal and difficult conditions.

a) Under **ordinal** conditions, movements towards the erroneous cabin (RC) were on average more than twice as fast as movements towards the correct cabin (GC), both in the LC and in summary: **a1**) in the LC: MT1(RC) = 22.5 < MT1(GC) = 46.2; p<0,01. **a2**) in summary: MT2(RC) = 43.7 < MT2(GC) = 102.7; p<0,01, (Table 1. Fig. 2, the left part).

b) On contrary, in **difficult** conditions movements towards the erroneous cabin (RC) were slower than movements towards the correct cabin (GC): **b1**) in the LC: MT3(RC) = 33.5 > MT3(GC) = 27.0; p<0,05. **b2**) in summary: MT4(RC) = 60.8 > MT4(GC) = 45.7 as a tendency (Table 1. Fig. 2, the right part). It agrees with the Swensson's rule established in humans initially.

II. Numbers and directions of snakes' head and/or body turns

Apparently, repeated turns of snakes' head and/or body to the right and to the left before correct and erroneous choices reflected their orientation, information collection and accumulation and possibly their hesitations during the pre-decision process before the final decision on a cabin choosing.

c) Under both ordinal and difficult conditions, numbers of turns was **greater** towards **a cabin chosen (CC) than for a cabin rejected (CR)**, both for correct choices and for erroneous ones. **c1**) Under **ordinal c**onditions: for **correct** choices: MNT1(CC) = 5,4 > MNT2(CR) = 3.9; p<0,01; for **erroneous** choices: MNT3(CC) = 3.9 > MNT4(CR) = 2.6; p<0,01 (Table 3; Fig. 3, the left part). **c2**) Under **difficult** conditions: for **correct** choices: MNT5(CC) = 6.7 > MNT6(CR) = 4.3; p<0,01; for **erroneous** choices: MNT7(CC) = 3.8 > MNT8(CR) = 2.7; p<0,01 (Table 2; Fig. 3, lower part). (Table 1; Fig. 3, the right part). These facts indicate a future choice preparation and verification during a pre-decision process.

d) Under both ordinal and difficult conditions, a number of turns was greater for correct choices than for erroneous ones both to cabins chosen and to cabins rejected. **d1**) Under **ordinal** conditions: **to a cabin chosen**: MNT1(GC) = 5.4 > MNT3(RC) = 3.9 as a tendency, **to a cabin rejected**: MNT2(GC) = 3.9 > MNT4(RC) = 2.6; p<0,05. **d2)** Under **difficult** conditions: to a **cabin chosen**: MNT5(CC) = 6.7 > MNT7(RC) = 3.8; p<0,05; to a **cabin rejected**: MNT6(GC) = 4.3 > MNT8(RC) = 2.7 as a tendency. (Table 1; the diagram looks just similar to the Fig. 3 and is not shown here therefore).

Thus, the snakes chose their way in the T-maze after the both ways comparing several times (in average, 2,39–3,83 turns to the GC and RC in each trial).

Probably, these turns served for information (evidences) reception in favor of response alternatives compared. Our important finding shows that a certain alternative (a cabin) is chosen for which there were more turns made. It corresponds to models of DM in humans which state that an alternative is chosen for which more evidences have been accumulated (Vickers, Lee, 1998). The notion of evidence was described in detail recently (Shendyapin, 2016).

Suggested calculation of numbers of animal's turns towards each of response alternative may serve as a method of indirect quantitative estimation of the evidences value. We know only one work (in humans) in which such values were assessed as numbers of flashes in 2 screens, 1 of the screens an observer had to choose (in which there were more flashes; Vickers, Pietsch, 2000).

Table 1. Average time values of snakes' staying in the T-maze

Ordinal conditions				Difficult conditions			
MT1 (RC)	MT1 (GC)	MT2 (RC)	MT2(GC)	MT3(RC)	MT3 (GC)	MT4 (RC)	MT4(GC)
22,5	46,2	43,7	102,7	33,5	27,0	60,8	45,7

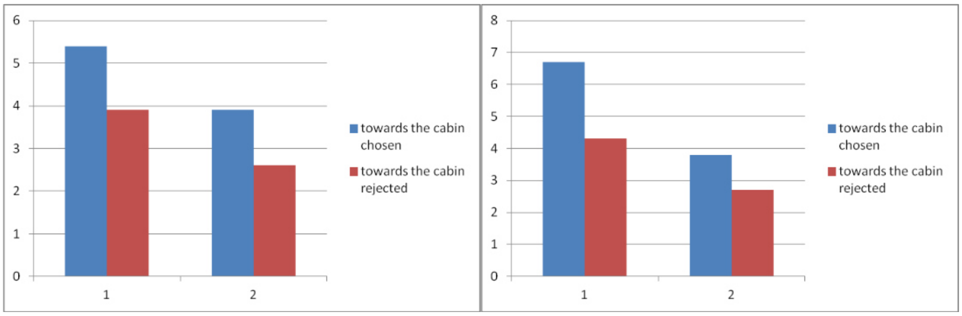


Fig. 2. Average values of time of snakes' staying in the T-maze under ordinal (the left part) and difficult (the right part) conditions (1 — in the LC, 2 — as a whole).

Table 2. Average values of numbers of turns in ordinal and difficult conditions

Ordinal conditions				Difficult conditions			
Correct choices		Erroneous choices		Correct choices		Erroneous choices	
MNT1 (CC)	MNT2 (CR)	MNT3 (CC)	MNT4 (CR)	MNT5 (CC)	MNT6 (CR)	MNT7 (CC)	MNT8 (CR)
5,4	3,9	3,9	2,6	6,7	4,3	3,8	2,7

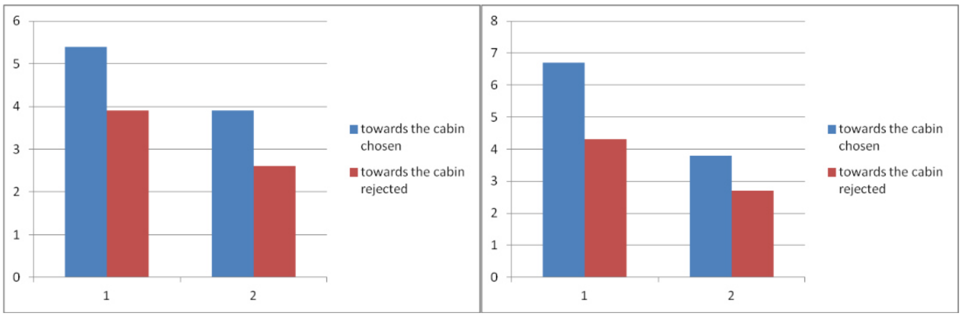


Fig. 3. Average numbers of turns towards the cabin chosen (CC) and towards the cabin rejected (CR) in ordinal (the left part) and in difficult (the right part) conditions (1 — correct choices, 2 — erroneous choices).

III. Individual behavior patterns

Individual data classification regarding medians of erroneous choices proportions (EP) and time of snakes' staying in the LC (as time of a choice (TC) probably) has shown the following (Fig. 4; there is no place for a big data table here, unfortunately). In the snakes 1 and 2, EP values (39 и 42) are less than the median of EP (Med_{EP} = 43) and TC values (24,6 и 25,8) are also less than the median of TC (Med_{TC} = 32.2) — so these snakes may be called "fast and accurate" conditionally. In the snake 3, EP value (37) is less than Med_{EP} while TC (38.1) is greater than Med_{TC} — it may be called "reflective" conditionally. Also, it has demonstrated the maximal number of turns as estimated for a single trial (3,83), that is similar to many information analysis operations made by reflective persons. In the snake 4, EP value (44) is almost equal to Med_{EP}, TC value (36.5) is greater than Med_{TC} and it occupies the 2nd place regarding a number of turns for a single trial (3,24), after the snake 3 ("reflective" exactly) who has shown 3,83 turns for a single trial. The snake 4 may be called "reflective" conditionally as well.

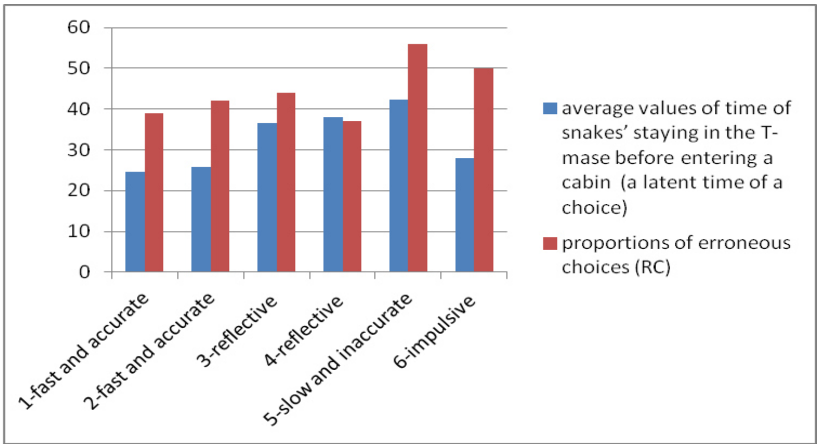


Fig. 4. Individual behavior patterns of 6 grass snakes.

In the snake 5, EP value (56) is grater than Med_{EP} and TC value (42,2) is greater than Med_{TC} — it may be called "slow and inaccurate" conditionally. And in snake 6, EP value is grater than Med_{EP} and equals to 50 (so it's choices look to be random), TC value (27,9) is less than Med_{TC} — it may be called "impulsive" conditionally.

Acknowledgements

The work was supported by Russian Foundation of Basic Research. Project № 18-013-00148.

References

Anokhin P.K. (1974). Biology and Neurophysiology of Conditioned Reflex and Its Role in Adaptive Behavior. 1st ed.; Pergamon Press: Oxford, UK.

Fetsch C. R., Kiani R. et al. (2014). Effects of Cortical Microstimulation on Confidence in a Perceptual Decision. *Neuron,* 83, 797–804, August 20.

Kagan J. (1966) Reflection–impulsivity: the generality of dynamics of conceptual tempo. *Journal of Abnormal Psychology*, 71, 17–24.

Kozeletsky Yu. (1977). Psychological theory of decisions. Warszava: PWN (in Polish).

Khvatov I.A., Sokolov A.Yu., Kharitonov A.N. (2015). Scheme of a body in snakes Lampropeltis triangulum campbelli. *Experimental Psychology*, 8, 119–138 (in Russian).

Lak A., Costa G. M. et al. (2014). Orbitofrontal Cortex Is Required for Optimal Waiting Based on Decision Confidence. *Neuron* 84, October 1, 90–201.

Ochinskaya E.I., Flyoss D.A. (1990). On an ability of mursh turtle to extrapolate a direction of a stimulus movement correctly. *Comparative Physiology of Higher Nervous Activity*, 106–113 (in Russian).

Safarov H.M. (1990) *Ecology and higher nervous activity in reptiles*. Dushanbe: Donish (in Russian).

Shendyapin V.M. (2016). Development of confidence model on the basis of Signal Detection Theory. In: I.G. Skotnikova et al. (Eds.) Fechner Day'2016. Proc. ISP 32 Ann. Meet. (p. 33). Moscow: Institute of Psychology Publishing House.

Smith P., Ratcliff R. (2004). Psychology and neurobiology of simple decisions, *Trends in neuroscience,* 27, 161–168.

Sokolov E.N. (2003) Perception and conditioned reflex: new look. Moscow: UMK "Psychology" (in Russian).

Swensson R.G. (1972). The elusive trade-off: Speed versus accuracy in visual discrimination tasks. *Perception and Psychophysics*, 12, 16–32.

Tversky A., Slovic P., Kaneman D. (1982). Judgment under in uncertainty: Heuristics and biases. Cambridge University Press.

Vickers D., Lee M (1998). Dynamic models of simple judgments: I. Properties of a self-regulating accumulator model. *Nonlinear dynamics, psychology and life sciences,* 2, 169–194.

Vickers D., Pietsch A. (2000). Decision making and memory: predicting accuracy, response time and confidence on individual trials. In: C. Bonnet (Ed.). Fechner Day'2000: Proc. ISP 10th Ann. Meet. (pp. 115–120). Strasbourg.

Zaichenko M.I., Sharkova A.V. et al. (2016). In extremly impulsive rats signal memory is manifested better than in low impulsive animals in 8-channels radial maze. *Journal of Higher Nervous Activity*, 66, 5, 1–11 (in Russian).

COMPARISON OF WEBER FRACTIONS OBTAINED IN A SIMULATION STUDY AND A BEHAVIORAL EXPERIMENT USING UNFORCED CHOICE PARADIGMS

Christian Hatzfeld[1], Mario Kupnik[1], and Florian Kattner[2]
[1] *Technische Universität Darmstadt, Measurement and Sensor Technology, Merckstr. 25, 64283 Darmstadt*
[2] *Technische Universität Darmstadt, Institute of Psychology, Alexanderstr. 10, 64283 Darmstadt*
hatzfeld@must.tu-darmstadt.de, kattner@psychologie.tu-darmstadt.de

Abstract

Unforced choice tasks (i.e. tasks that allow a "don't know"-response) are expected to reduce response uncertainty for test subjects, especially at low stimulus intensities. In this work, we present simulated and behavioral data for unforced choice tasks using standard staircase (2-down/1-up), and parametric (updated maximum likelihood; UML, and Bayesian adaptive estimation; Psi) procedures. Monte Carlo simulations were conducted for Yes-No and 3-interval-forced-choice tasks with and without the option to give "don't know" responses. In the experiment, Weber fractions of a vibrating haptic stylus (signal frequency 25 Hz, reference amplitude 40 µm (peak-peak)) were obtained from 12 participants (direct comparison of test and reference stimulus) for each procedure. Both simulation and experiment contained 40 trials. In the experiment, average Weber fractions differed significantly between forced and unforced-choice tasks. Simulations underestimate the variation coefficient in almost all conditions, with higher values for unforced-choice tasks regardless of the procedure. Overall, simulation results could not be reproduced in the experiment.

Introduction

Implementation of unforced-choice tasks

Kaernbach introduced a two-step decision model for unforced choice tasks (Kaernbach, 2001, nAUC – n alternative unforced choice tasks in the following). In the first step, the participant decides whether to respond to the stimulus or give an "I don't know" response. This is modeled by a so-called safety margin δ, that is added to the chance performance of the current task. Only if the assumed probability to score a correct response is greater than this value, a response will be provided by the participant. Kaernbach (2001) describes different options to model the assumed probability, an algorithmic approach is given by the authors (Hatzfeld et al., 2015).

Overall, these considerations lead to probabilities for correct, false and unsure responses for each task. Based on these probabilities, the progression rules of existing psychometric procedures can be extended to incorporate unforced choice tasks. This was done for three procedures described as follows:

- For the well-known *staircase procedure* (Cornsweet, 1962; Levitt, 1971), an additional step size is defined in the same way as described by Kaernbach

(2001), i.e. modeling the probability correct of repeated guesses. The step size after an "unsure" response (s_{unsure}) is calculated based on the convergence probability of the staircase (p_{equ}) and the step sizes after false and correct responses (s_{false} and s_{correct}, respectively) according to

$$\frac{s_{\text{false}}}{-s_{\text{correct}}} = \frac{p_{\text{equ}}}{1-p_{\text{equ}}} \quad \text{and} \quad \frac{s_{\text{unsure}}}{-s_{\text{correct}}} = \frac{p_{\text{equ}}-1/n}{1-p_{\text{equ}}} \,. \qquad (1)$$

- Parametric procedures such as Psi (Kontsevich & Tyler, 1999) and UML (Shen & Richards, 2012) are operating on a set of candidate functions and calculate the most likely candidate based on the probabilities of correct and false responses and the course of the experiment. By including the option of an unsure response to the calculation of the most likely candidate function, unsure responses can be included fairly easy in parametric procedures. This was done by expanding the entropy calculation for the Psi procedure as well as the maximum likelihood calculation for the UML procedure (as described in Hatzfeld et al., 2015).

Methods

Simulation of unforced-choice Tasks

In order to gain insight into the performance of the procedure, a Monte Carlo simulation was carried out. This simulation is based on a psychometric function according to

$$p(x) = \gamma + (1-\gamma-\lambda)\cdot\frac{1}{1+e^{-\beta(x-\alpha)}}. \qquad (2)$$

with lapse rate $\lambda=0.01$, and the guess rate γ determined from the employed task (i.e., $\gamma = 0$ for YN/YUN tasks and $\gamma = 0.33$ for 3AFC/3AUC tasks). Threshold α and slope β were calculated from the behavioral data for each combination of task and procedure. For the simulation, the above described two-step decision model is implemented with a safety margin of $\delta = 0.1$, such comprising a model of a participant in the psychophysical experiment.

Monte-Carlo simulation was carried out in Matlab (version 2014b, The MathWorks, Natick, MA, USA) with 3000 experimental runs and the mean threshold as well as the standard deviation was recorded.

Acquisition of Behavioral Data

We measured Weber fractions for a haptic stimulus with a frequency of 25 Hz and a reference intensity of 40µm (peak to peak deflection). Stimuli were delivered with a stylus (Ø 12 mm, Figure 1a). The stylus was driven by a commercial mini shaker (model 4810, Bruel & Kjaer, Naerum, DK) driven by an analogue linear amplifier (model 2760, Bruel & Kjaer). A closed-loop servo control implemented on a FPGA board (model 7833R, National Instruments, Austin, TX, USA) with an acceleration sensor of an impedance head (model 8001, Bruel & Kjaer) was used to control the shaker system.

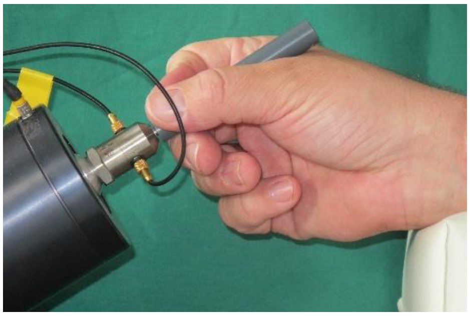
(a) 
(b) 

Figure 1. Experimental setup. (a) Stimulus delivery, (b) response console indicating the possible responses for the ongoing experiment (in German language). In this case, a YUN task (Yes – *Ja*, No – *Nein*, Don't know - *??*) is displayed.

Participants gave their responses on a console (Figure 1b) based on an Arduino micro controller. Providing a display and four dedicated buttons, this setup allows to display the possible responses in the ongoing task. Furthermore, response times can be recorded by the console. The console is controlled over an USB serial connection by the superordinate experimental control software implemented in LabView (version 2014, National Instruments).

48 participants were recruited at Technische Universität Darmstadt for the experiment. They were randomly assigned to the four different experimental groups as summarized in Table 1. Each participant performed a test run with 15 trials to familiarize with the stimuli and response input device. Subsequently, they performed 40 trials for each experimental condition. Conditions were presented in random order. Participants gave informed consent before the experiment, the experimental protocol was approved by the ethics board of Technische Universität Darmstadt.

Parameters of the psychometric function were fitted from the experimental data.

Table 1. Experimental groups, experimental conditions were presented in random order. Test run data are not included in the data analysis.

Group	**Participants**		**Performed Tasks**				
	Age	Gender	Test Run	Experimental conditions			
1 *UML*	23.5 ± 2.8	6m, 6w	SC3 YN	UML YN	UML YUN	UML 3AFC	UML 3AUC
2 *Staircase*	22.9 ±3.4	3m, 9w	SC3 YN	SC2 YN	SC2 YUN	SC2 3AFC	SC2 3AUC
3 *Psi*	22.9 ±3.6	6m, 6f	SC3 YN	Psi YN	Psi YUN	Psi 3AFC	Psi 3AUC

Abbreviations: SC2 – Staircase procedure with 2down1up progression rule, SC3 – Staircase procedure with 3down1up progression rule.

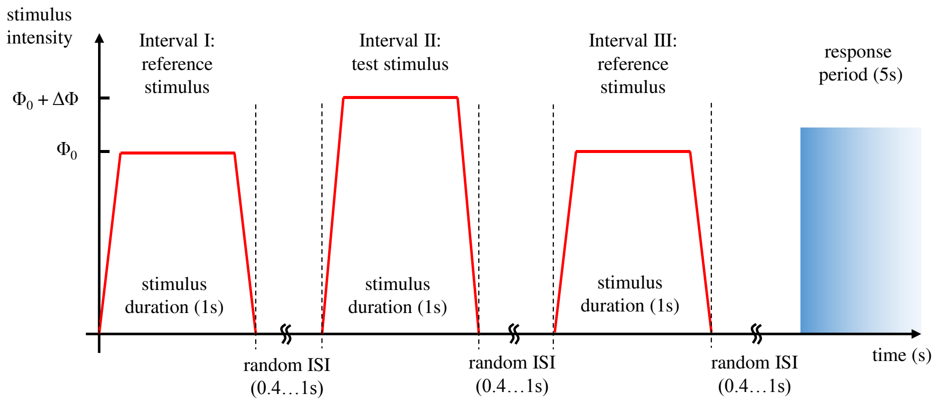


Figure 2. Example of stimulus presentation for a 3-alternative condition: two identical test stimuli and one test stimulus are presented in random order with randomly varying ISI (Inter-stimulus-intervals). Stimuli have a 100 ms linear on- and offset to prevent high-frequency signal components. In the response period; participants have to indicate the interval with the test stimulus (Interval II in this example). For Yes-No-tasks, only two intervals are used, and participants have to indicate, whether they feel a difference between the two stimuli.

Results

We calculated the Weber fractions according to the calculation rules defined by the procedures. Table 2 provides the results for each experimental condition as Weber fraction over all participants. For better comparison of behavioral and simulated results, standard deviations were normalized with respect to the mean.

Table 2. Comparison of behavioral and simulated data. No correction was made for different convergence levels of the staircase methods.

		Experiment		**Simulation**		
Procedure	Task	Weber Fraction	variation coefficient (std/mean)	Weber Fraction	variation coefficient (std/mean)	Bias (%)
Staircase (2down-1up)	YN	1,128	0.091	1,518	0,149	34,6%
	YUN	0,995	0.107	1,401	0,139	40,8%
	3AFC	1,066	0.074	1,233	0,073	15,6%
	3AUC	1,077	0.130	1,208	0,069	12,2%
Psi	YN	1,361	0.223	1,351	0,292	-0,8%
	YUN	1,266	0.258	1,172	0,046	-7,4%
	3AFC	0,755	0.048	1,466	0,134	94,0%
	3AUC	1,093	0.198	1,054	0,029	-3,6%
UML	YN	1,346	0.139	1,371	0,096	1,9%
	YUN	1,093	0.264	1,178	0,084	7,8%
	3AFC	1,053	0.059	1,164	0,314	10,5%
	3AUC	1,245	0.181	2,004	0,649	61,0%

A 3 (procedure: Staircase, Psi, UML) × 4 (task: YN, YUN, 3AFC, 3AUC) repeated-measures ANOVA of the Weber fraction reveals significant main effects for procedure ($F_{(2,33)} = 6.82$, $p = .003$, $\eta_G^2 = .13$)[1] and task ($F_{(3,99)} = 47.41$, $p < .001$, $\eta_G^2 = .47$), as well as a significant interaction ($F_{(6,99)} = 8.92$, $p < .001$, $\eta_G^2 = .25$).

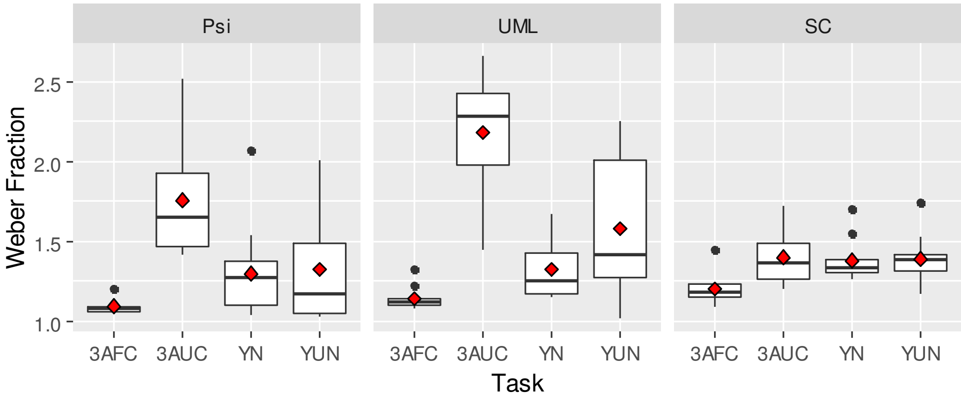


Figure 3. Boxplot of behavioural results. Red rhombus indicates the arithmetic mean of the data.

Discussion

Experimental results show mostly reasonable values for a Weber fraction, i.e. values larger than 1. Unexpectedly, results from a staircase procedure with YN- and YUN-task are smaller than values from parametric procedures, although the convergence level of staircase procedures is normally higher than that of parametric procedures. Possible reasons include a tendency to yes-responses of the participants but were not investigated further.

The comparison of sample means from behavioral and simulated data shows no significant difference in about half of the experimental conditions (Tab. 2, bias column). This contradicts studies such as [Madigan1987, Karmali2016], where simulation results resemble experimental results more closely. Reasons for this difference could be different procedures considered in the different studies as well as other experimental conditions (progression and placement rules, experiment duration). Smaller differences (lower bias) between experimental and simulated data can be found for parametric procedures. Especially parametric procedures such as Psi and UML with a YN- ask exhibit low mean differences between experiment and simulation. Larger differences exist for non-parametric procedures as well as forced choice tasks.

[1] Effect sizes were calculated as generalized eta squared according to (Bakeman, 2005), p-values were GG-adjusted.

Conclusion

This work does not give a clear indication, whether the usage of an unforced choice task will yield more accurate results. For each procedure, there are combinations with forced choice and unforced choice paradigms, that converge at similar nominal values.

On the other hand, other praised advantages of unforced choice tasks have to be investigated further. This includes the perceived complexity and workload of the participant and will be focus of our further work.

Acknowledgements

The authors thank Stefan Gries and Knut Drewing for helpful discussions regarding the experimental design and Tobias Wähner for conducting the experiments.

References

Cornsweet, T.: The Staircase-Method in Psychophysics. American Journal of Psychology, 1962, vol. 75, pp. 485-491.

Hatzfeld, C, Kupnik, M & Werthschützky, R.: Performance Simulation of Unforced Choice Paradigms in Parametric Psychometric Procedures. Worldhaptics Conference 2015, Evanston, IL, USA.

Kaernbach, C.: Adaptive Threshold Estimation with Unforced-Choice Tasks Attention, Perception & Psychophysics, 2001, vol. 63, pp. 1377-1388.

Kontsevich, L. & Tyler, C.: Bayesian Adaptive Estimation of Psychometric Slope and Threshold. Vision Research, 1999, vol. 39, pp. 2729-2737.

Levitt, H.: Transformed Up-Down Methods in Psychoacoustics. Journal of the Acoustical Society of America, 1971, vol. 49, pp. 467-477.

Shen, Y. & Richards, V. M.: A maximum-likelihood procedure for estimating psychometric functions: Thresholds, slopes, and lapses of attention. Journal of the Acoustical Society of America, 2012, vol. 132, pp. 957-967.

R. Madigan, D. Williams, Maximum-likelihood psychometric procedures in two-alternative forced-choice: Evaluation and recommendations, Attention, Perception, & Psychophysics 42(3), 240 (1987). DOI 10.3758/BF03203075

F. Karmali, S.E. Chaudhuri, Y. Yi, D.M. Merfeld, Determining thresholds using adaptive procedures and psychometric fits: evaluating efficiency using theory, simulations, and human experiments, Experimental brain research 234(3), 773 (2016).

FECHNER'S THREE EXPERIMENTAL METHODS: DO THE RESULTS CORRELATE?

Nan Yang[1,2], Roger D. Adams[1], Jia Han[2], Gordon Waddington[1]
[1] *Faculty of Health, University of Canberra, Bruce, ACT 2617, Australia*
[2] *Shanghai University of Sport, Yangpu Qu, Shanghai Shi, China*
nan.yang@canberra.edu.au; roger.adams@canberra.edu.au;
gordon.waddington@canberra.edu.au; jia.han@canberra.edu.au

Abstract

Fechner in 1860 outlined three experimental methods – limits, adjustment, and right and wrong cases, the three fundamental psychophysical methods that Boring noted were still in constant use one hundred years later. Each of these methods was proposed as a way of determining a just noticeable difference that could be either absolute or differential. Although the definition of a correlation did not appear until Galton's paper in 1888, and the computational formula until Pearson's paper in 1895, it might be assumed that Fechner believed that results from the three methods would be correlated. When testing acuity of proprioception, the perception of information about movement of the body and body segments, it is possible to apply the methods with the same participants, obtain sensitivity scores, and determine their inter-correlation. In this study, the method of adjustment and a signal detection theory-based discrimination task were employed with the same apparatus and same participants to derive sensitivity indices for the extent of ankle inversion movements. The scores obtained by the two methods were not significantly correlated. A search of the literature revealed a similar lack of correlation obtained between the JND for active inversion movements (constant stimuli) and detection of imposed passive inversion movements (limits) at the ankle, and similarly uncorrelated sensitivity indices for the method of adjustment and method of limits applied in tests of proprioceptive function at the knee joint. Thus, for proprioception testing, no pair-wise combination of Fechner's three methods shows a significant correlation between the obtained sensitivity scores. It is possible that the observed lack of correlation between the measures derived from the three different methods emerges as a result of different cognitive processes involved in task performance.

Judging lifted weights without visual information, a task based on proprioception, was what Fechner made 'the representative psychophysical experiment' (Boring, 1942). Later, in his History of Experimental Psychology, Boring (1950) observed that Fechner's claim to greatness within psychology was that he 'conceived, developed and established three new methods of measurement…that are still used with only minor modifications'…the methods of limits, right and wrong cases (later constant stimuli), and average error (later called the method of adjustment or the method of reproduction). Each of these methods was proposed as a way of determining sensitivity on a sensory continuum. However, correlation was first defined by Galton (1888), and the computational formula provided by Pearson (1895) so the idea of statistically examining the relationship between the measures would not have occurred to a nineteenth-century psychophysicist.

The focus of the present work is on how the three methods - limits, adjustment, and constant stimuli - are applied in testing proprioception, and how the scores derived from them inter-relate. Which of the three methods is most appropriate for testing joint proprioception continues to be a matter of debate, with each of the methods being supported in recent reviews. Goble (2010) has argued for the ‘relative efficiency’ of joint position reproduction (JPR), given that it was ‘generally well-accepted that matching errors can be a useful indicator of proprioceptive acuity’ (p.1177), with individuals who make large position matching errors being held to be ‘proprioceptively deficient’. Elangovan et al (2014) proposed that threshold to detection of passive motion (TTDPM) testing better reflects afferent proprioceptive feedback processing ability, because joint position matching methods typically require active motion. According to these researchers, testing detection of passively-induced joint motion gives the ‘purest’ measure of proprioceptive function. Lastly, Han et al (2015) have emphasized the ‘ecological validity’ of the active movement extent discrimination assessment (AMEDA) method (Waddington & Adams, 1999), which involves testing joint proprioceptive acuity with a constant set of stimuli (movements made to physical stops, with different extents created by different end positions), either in comparison with a standard movement on each trial, or by the method of single stimuli (Woodworth & Schlosberg, 1954). Good scores are related to achievement, in that ankle proprioception tested in this way is significantly correlated with attained level of competitive sports performance (Han et al., 2015).

From the different methods, low reproduction error in JPR, a low threshold in TTDPM, and a small jnd with constant stimuli are all taken to represent good proprioceptive sensitivity, and should thus be positively correlated, whereas a larger AUC score reflects more acute discrimination, and negative correlations with other measures would be expected.

The aim of this study was to compare the AMEDA discrimination method and the JPR method for determining ankle proprioceptive sensitivity.

Method

Participants

Twenty participants with mean age 21.4 years volunteered. Self-report instruments showed all to be right-handed and right-footed (Yang et al, 2018). Each participant was evaluated using two separate ankle proprioception tests; the method of active movement extent discrimination, and in the other session, active reproduction of an actively-determined ankle joint inversion position. The order of the two tests was randomised, with a two-hour rest interval between.

Equipment

The same apparatus was used for both Joint Position Reproduction and Active Movement Extent Discrimination testing. An electrogoniometer on the footplate axle enabled the measurement of the reproduced inversion angle (Figure 1).

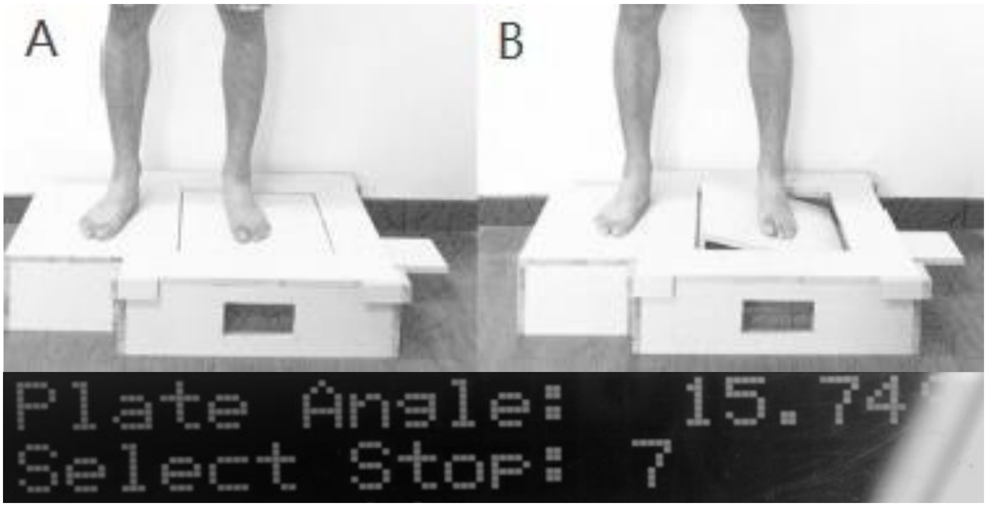


Figure 1. Apparatus with plate at horizontal (A), and after an ankle inversion movement to 15.74° (B). In the reproduction part of the JPR method, the plate movement into inversion was not physically stopped (7 was code for no stop), and the angle reading was recorded.

Active Movement Extent Discrimination Assessment

The Ankle Active Movement Extent Discrimination Assessment system (AMEDA) was used to test ankle inversion movement proprioception in full weight-bearing on a fixed wooden platform housing a square wooden plate that swung on an axle aligned to the long axis of the foot. For testing, participants looked straight ahead, and stood in an even weight-bearing stance, with one foot on the fixed platform and the other centred over the axle of the swinging plate (Fig. 1). On each trial, an active ankle inversion movement was made that rotated the swing plate from the horizontal start position downwards until the plate edge contacted the height-adjustable stop. Different movements were employed, in this case five inversion extents: 10°, 11°, 12°, 13°, and 14°. Before the test, participants were given a familiarisation session with the four ankle inversion positions in order, from the smallest through to the largest (position 4=16°), 3 times. During the test, participants undertook 50 trials without feedback, with 10 randomly-presented trials for each of the 5 different inversion displacements. Participants were asked to identify the different ankle inversion positions by responding with a number (1, 2, 3, 4, or 5) indicating the ankle inversion position that they felt they had just experienced.

Active Joint Position Reproduction Assessment

The same apparatus was used to assess ankle inversion movement proprioception by the method of active joint position reproduction, through the addition of an electrogoniometer to continuously record plate angle. In order to maintain a similar protocol to the active movement extent discrimination test, the JPR method of an active criterion movement with an active ipsilateral reproduction movement was employed. During testing, the computer-controlled stepper motor set the height-adjustable stop for a selected inversion movement extent and the participant actively made an inversion movement till the swinging plate contacted the physical stop, actively returning the plate to horizontal after three seconds. When the height-adjustable stop was withdrawn, the participant was asked to actively reproduce the previously experienced position, verbally indicating when they were satisfied the position was an accurate reproduction of the criterion movement stopped end position. Five angles of 10°,11°, 12°, 13°, and 14° were used in JPR testing, each randomly presented and reproduced ten times. Proprioceptive ability was measured by the mean Absolute Error (AE) of the difference in degrees between the angle presented and the angle reproduced.

Data analysis

For the JPR tests, absolute error (AE) was employed to represent the participant's proprioceptive acuity. The absolute error (AE) between the actual degree of the target position and the position reproduced by the participant was calculated and then the mean of the AEs for the 50 trials was obtained to represent the participant's proprioceptive performance in the test. For the AMEDA tests, pair-wise Receiver Operating Characteristic (ROC) curves were produced by non-parametric signal detection analysis, comparing responses to distances 1 and 2, 2 and 3, 3 and 4, 4 and 5 (Adams et al., 2017). The mean of the pair-wise Area Under the Curve (AUC) scores was calculated to obtain a single ankle movement discrimination score for each participant, representing the participant's proprioceptive acuity for discriminating pairs of movement extents in the 10° – 14° range. The relationship between the results of the two methods was assessed using Pearson's product-moment correlation. Significance was set at p less than 0.05 throughout. All statistical analyses were carried out using the Statistical Package for the Social Sciences (SPSS, version 21.0).

Results

For the twenty participants, AMEDA Mean AUC = 0.68, SD = 0.06, JPR Mean Absolute Error =1.68°, SD = 0.51°. The correlation between the AUC ankle movement discrimination scores and the JPR ankle movement reproduction error scores was $r = -0.26$, $p = 0.17$. A search of the literature revealed five other studies that compared two of Fechner's three methods for assessing joint proprioception (Fig. 2). Three of these studies worked with movements at the knee joint, Nagai et al. (2016), Grob et al (2002) and Li et al (2016), and all compared JPR and TTDPM methods, reporting correlations between scores of 0.22, 0.16 and 0.29 respectively, with all $p > 0.12$. The two studies that compared a theshold from TTDPM testing with a discrimination score, De Jong et al (2005) and Elangovan et al (2014), both calculated a jnd discrimination score, from application of the method of constant stimuli by De Jong et al, and from Elangovan et al's use of an adaptive algorithm that selected the value of the comparison elbow flexion stimulus on the subsequent trial from the previous response, to ensure monotonic convergence on the joint position difference threshold. Neither correlation coefficient, -0.20 and -0.08, was significant, with both $p > 0.43$.

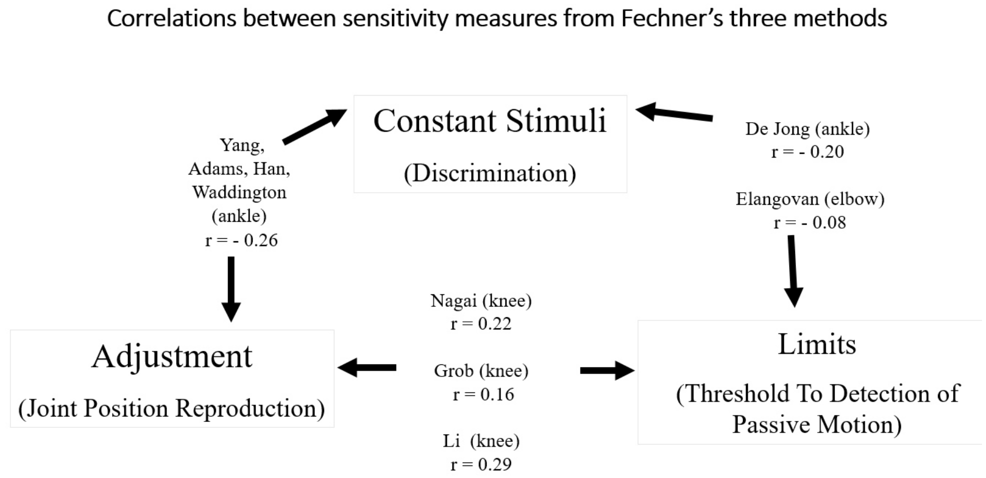


Figure 2. Correlations between the proprioceptive sensitivity measures obtained from Fechner's three methods in six studies.

Discussion

The current data complete the last segment of the triangle of pair-wise comparisons between Fechner's three assessment methods for proprioceptive sensitivity, and the correlations between methods show that the measures arising are test-specific. This adds a test-specific feature of proprioceptive sensitivity measures to their site-specific nature, given the previous Han et al (2013) finding that using the same method (constant stimuli), the obtained AMEDA scores are site-specific, with the mean correlation between proprioceptive performance on the left and right sides at the ankles, knees, shoulders and hands being 0.9, and the mean correlation between proprioceptive performance at different joints on the same side being 0.1.

Indeed, Krewer et al (2015) suggested that the three methods test different aspects of the proprioceptive sense, are not interchangeable, and as such, intercorrelation would not be expected. Differences in the memory requirements of the different tasks may be the cause (Elangovan et al., 2014). Consideration of the separate task requirements reveals these differences: detection of a passively-applied joint motion; holding the criterion movement extent and location in working memory long enough to make the reproduction movement; and judging a given active movement to a stop as one of a fixed set that has previously been established during familiarization. These are tasks that vary in memory demand, most obviously in terms of the 'depth of processing' of the proprioceptive information (Craik & Lockhart, 1972). For the present, then, it is likely that researchers will use the measures that they determine to have been validated by showing sensitivity to different levels on independent variables of interest, e.g. the dimensions of injury, disease, and sport expertise.

Conclusion

For tests of proprioceptive sensitivity at the same joint, Fechner's three methods give sensitivity scores that are not correlated between the three different testing techniques.

References

Adams, R.D., Waddington, G., Steinberg, N. and Han, J. (2018) John Swets, ROC curves, and fifty years of Signal Detection Theory and Psychophysics. Part VII, *Proceedings of Fechner Day*, *33*(1).

Boring, E.G. (1950) *History of experimental psychology*. London: Prentice-Hall International.

Boring, E.G. (1942) *Sensation and perception in the history of experimental psychology*. New York: Appelton-Century.

Craik, F.I. and Lockhart, R.S. (1972) Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), pp.671-684.

de Jong, A., Kilbreath, S.L., Refshauge, K.M. and Adams, R. (2005) Performance in different proprioceptive tests does not correlate in ankles with recurrent sprain. *Archives of Physical Medicine and Rehabilitation*, *86*(11), pp.2101-2105.

Elangovan, N., Herrmann, A. and Konczak, J. (2014) Assessing proprioceptive function: evaluating joint position matching methods against psychophysical thresholds. *Physical Therapy*, *94*(4), pp.553-561.

Fechner, G. (1860/1966) *Elements of Psychophysics*. New York; Holt, Rinehart & Winston.

Galton, F. (1889) Co-relations and their measurement, chiefly from anthropometric data. *Proceedings of the Royal Society of London*, *45*(273-279), pp.135-145.

Goble, D.J. (2016) Proprioceptive acuity assessment via joint position matching: from basic science to general practice. *Physical Therapy*, *90*(8), pp.1176-1184.

Grob, K.R., Kuster, M.S., Higgins, S.A., Lloyd, D.G., and Yata, H. (2000) Lack of correlation between different proprioceptive movements in the knee. *Journal of Bone and Joint Surgery*, 84-B, 614-618.

Han, J., Anson, J., Waddington, G. and Adams, R. (2013) Proprioceptive performance of bilateral upper and lower limb joints: side-general and site-specific effects. *Experimental Brain Research*, *226*(3), pp.313-323.

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.

Han, J., Waddington, G., Adams, R., Anson, J. and Liu, Y. (2016) Assessing proprioception: a critical review of methods. *Journal of Sport and Health Science*, *5*(1), pp.80-90.

Krewer, C., Van de Winckel, A., Elangovan, N., Aman, J.E. and Konczak, J. (2016) Commentary on:"Assessing proprioception: A critical review of methods" by Han et al. *Journal of Sport and Health Science*, *5*(1), pp.91-92.

Li, L., Ji, Z.Q., Li, Y.X. and Liu, W.T. (2016) Correlation study of knee joint proprioception test results using common test methods. *Journal of Physical Therapy Science*, *28*(2), pp.478-482.

Nagai, T., Allison, K.F., Schmitz, J.L., Sell, T.C. and Lephart, S.M. (2016) Conscious proprioception assessments in sports medicine: how individuals perform each submodality?. In *Sports Medicine* (pp. 1-13). SM Online Scientific Resources, Dover, DE.

Pearson, K. (1896) Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character*, *187*, pp.253-318.

Waddington, G. and Adams, R. (1999) Discrimination of active plantarflexion and inversion movements after ankle injury. *Australian Journal of Physiotherapy*, *45*(1), pp.7-13.

Woodworth, R.S. and Schlosberg, H. (1954) *Experimental Psychology*. Oxford and IBH Publishing.

Yang, N., Waddington, G., Adams, R. and Han, J. (2018) Translation, cultural adaption, and test–retest reliability of Chinese versions of the Edinburgh Handedness Inventory and Waterloo Footedness Questionnaire. *Laterality: Asymmetries of Body, Brain and Cognition*, *23*(3), pp.255-273.

SESSION VII

TWO-DIMENSIONAL SCALING OF ENVIRONMENTAL ODORS IN FIELD-STUDIES

Kirsten Sucker[1], Ralf Both[2], Frank Müller[3], Michael Bischoff[3], and Gerhard Winneke[4]
[1] *Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr-University Bochum (IPA), D-44789 Bochum, Germany*
[2] *North Rhine-Westphalia State Agency for Nature, Environment and Consumer Protection, Leibnizstraße 10, D-45659 Recklinghausen, Germany*
[3] *uppenkamp und partner, Kapellenweg 8, D-48683 Ahaus, Germany*
[4] *Medical Institute of Environmental Hygiene at Heinrich-Heine-University Düsseldorf, Auf'm Hennekamp 50, D-40225 Düsseldorf, Germany (retired)*
sucker@ipa-dguv.de

Abstract

Whereas laboratory applications of the Semantic Differential (SD) for describing affective dimensions of odors in laboratory settings are available, we used the SD for the description of odorous emissions from industrial or agricultural sources in field settings. Six industrial and three agricultural odors, with a priori ratings on the pleasantness-unpleasantness continuum, were chosen. Trained observers described the perceived odors using the SD, and affective content was ascertained by comparing individual odor-profiles with those of the mental concepts of "stench" and "fragrance". Within the industrial settings profile correlations were significantly positive for "fragrance" and negative for "stench" for rusk and sweets production, whereas for cast iron production and fat treatment correlations were the other way round. Within the agricultural settings all correlations were significantly positive for "stench" and negative for "fragrance". Principal Components Analysis (PCA) revealed two separate clusters within a two-dimensional structure, namely a stench- and a fragrance-cluster. These findings validated and improved a priori classifications of pleasantness-unpleasantness of environmental odors.

Introduction

The Semantic Differential (SD) or Polarity Profile (PP) was originally developed for the measurement of meaning within psycholinguistics (Osgood et al., 1957). It consists of n bipolar scales with opposite adjectives at the endpoints with descriptive (e.g. strong … weak) or affective content (e.g. good … bad). Connecting mean values on the n bipolar scales creates a profile. An illustrative example comparing profiles for perceived odors, namely thiophenol as a malodor and peppermint as a more pleasant odor, with the mental concept of "stench" is given in Figure 1. Profile similarity for thiophenol with the mental concept of "stench" (left part), quantitatively given by means of the profile correlation, is highly positive ($r = 0.98$), whereas for peppermint (right part) it is moderately negative ($r = -0.65$).

Within environmental psychology the SD has received wide applications for describing affective dimensions of e.g. buildings, landscapes, colors or sounds.

The semantic space of the SD has repeatedly been found to be essentially three-dimensional with axes evaluation, potency and activity (EPA-structure). Whereas laboratory applications for describing affective dimensions of odors in controlled laboratory settings are available (e.g. Dalton et al., 2007) we, to the best of our knowledge, for the first time used the SD for the description and evaluation of odorous emissions from industrial and agricultural facilities in field settings.

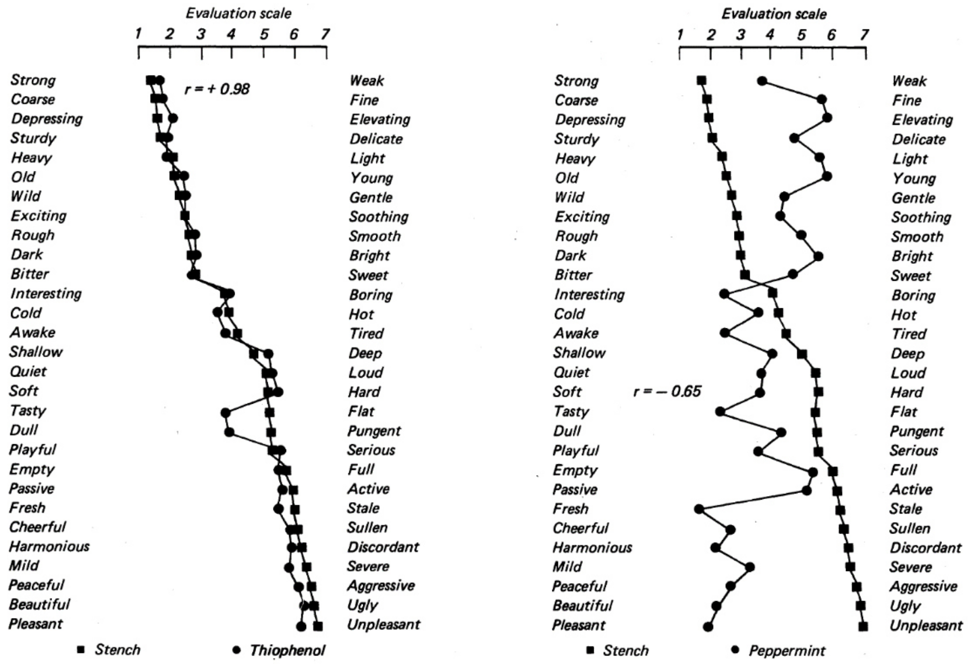


Figure 1. Polarity profiles from Guideline VDI 3882/2 (1994): Mental concept “stench” as compared with odor stimuli thiophenol (left) or peppermint (right) based on 29 pairs of adjectives. Correlation coefficients *r* are highly significant.

Method

Six industrial and three agricultural facilities with a priori ratings on the pleasantness-unpleasantness-continuum, as given by the Environmental Agencies of the participating German States, were chosen. 10 to 36 (industry) or 31 to 55 (agriculture) trained observers described the perceived odors downwind in the closest possible vicinity of the respective facility using the SD given in Figure 1 (randomized order of adjective descriptors), and affective content was ascertained by comparing individual odor-profiles with those of the mental concepts of “stench” and “fragrance” by means of product-moment correlation coefficients. The selection of observers was based on odor threshold determinations for n-butanol and hydrogen sulfide.

Results and Discussion

Profile correlations between the mental concepts of “stench” and “fragrance” and environmental odors from industrial and agricultural sources are given in Table 1. Within the odors in ambient air from industrial sources (A) only those from rusk and sweets production exhibit significant positive correlations with "fragrance" and negative correlations with "stench", whereas for the four remaining odors the correlation structure is the other way round. As for the odors in ambient air from the different agricultural facilities correlations with "fragrance" are negative throughout and positive for "stench".

Table 1. Profile correlations (plus levels of significance) between profiles describing environmental odors in ambient air from industrial (A) and agricultural (B) sources and profiles describing mental concepts of "fragrance" and "stench", respectively (n = 29).

	Fragrance		**Stench**	
	r	**p**	***r***	**p**
Industrial Sources (A)				
Sweets production	0.73	<0.001	- 0.52	0.004
Rusk production	0.84	<0.001	- 0.65	<0.001
Seed-Oil mill	- 0.18	0.36	0.51	0.005
Textile colouring	- 0.60	0.001	0.65	<0.001
Cast iron factory	- 0.82	<0.001	0.97	<0.001
Fat	- 0.75	<0.001	0.90	<0.001
Agricultural Sources (B)				
Poultry	- 0.81	<0.001	0.96	<0.001
Pig	- 0.74	<0.001	0.94	<0.001
Cattle	- 0.50	0.006	0.78	<0.001

Principal Components Analysis (PCA) of these associations enriched with additional stench- and fragrance-profiles revealed an orthogonal 2-dimensional structure within which two widely separated clusters characterized by pleasant and unpleasant environmental odors could clearly be identified.

These findings clearly confirmed a priori-classifications of different environmental odors proposed by the respective Environmental Agencies, but added more precise quantitative information to these largely binary classifications. Thus, they were instrumental in planning and conducting population-based annoyance studies in the vicinity of the different odor sources (Sucker et al., 2008). These studies clearly showed that only aversive and neutral environmental odors induce increased odor annoyance responses in a dose-related manner, whereas pleasant odors like those from rusk or sweets production do not. "Dose" in this context refers to the frequency of odor events within one year, as ascertained by systematic field observations or dispersion modelling.

Based on these findings, regulation of environmental odors in Germany was developed within the "German Guideline on Odor in Ambient Air" (GOAA, 2008) taking the pleasantness-unpleasantness dimension into account (VDI 3940/4, 2010).

Acknowledgements

These studies were partly supported by the German Chemical Industry Association (VCI), the Ministry of Environment, Climate Protection and Energy of Baden-Württemberg, and the Ministry of Environment, Energy and Climate Protection, State of Lower Saxony, Additional support was given by the Ministry of Agriculture and Environment of the State of Mecklenburg-Vorpommern, the Ministry for Environment, Agriculture, Conservation and Consumer Protection of the State of North Rhine-Westphalia, and the Saxonian Ministry of Environment and Agriculture.

References

Dalton P, Maule C, Oshida A, Hikichi S, Izumi Y (2008). The use of semantic differential scaling to define the multidimensional representation of odors. J Sens Stud. 23, 485 – 497.

Guideline on Odour in Ambient Air (GOAA 2008). Determination and Evaluation of Odour Immissions (2008). https://www.lanuv.nrw.de/fileadmin/lanuv/luft/gerueche/pdf/GOAA10Sept08.pdf (in English)

Guideline VDI 3882/Part 2 (1994). Olfactometry – Determination of Hedonic Odour Tone. Berlin: Beuth Verlag.

Guideline VDI 3940/Part 4 (2010). Determination of hedonic odour tone – Polarity profiles. Berlin: Beuth Verlag.

Osgood, C.E., Suci, G. Tannenbaum P (1957). The measurement of meaning. Urbana: University of Illinois Press.

Sucker K, Both R, Bischoff M, Guski R, Krämer U, Winneke G (2008). Odor frequency and odor annoyance. Part II: Dose-response associations and their modification by hedonic tone. Int Arch Occup Environ Health 81, 683 – 694.

SENSORY PERFORMANCE AND DECISION MAKING INDICES IN CHRONIC PATIENTS

T.A. Zabrodina and A.S. Sorokin
Tula State University (Tula)
tatzab53@mail.ru

The aim of the research was to study sensory performance in patients having several pathologies (oncology, cerebral palsy, allergic dermatosis). It was found that the discriminability level of patients depended on the type of pathology and could be either higher or lower than that in healthy participants. While the signal detection probability (Hits) was considerably lower in the cerebral palsy group than that in the healthy one, it was significantly higher in the allergic dermatosis group compared to the healthy group. The false alarm probability was much higher in all chronic disease groups than in healthy participants. The response time was greater in oncology patients and in patients with cerebral palsy especially. The decision-making criterion in chronic patients was more "liberal" than that in healthy participants (especially in allergic dermatosis group).

Change in a decision-making strategy may be due to the powerful non-sensory disease factor, which influences a sensory task performance and decreases a responsibility for it.

Studies of sensory and perceptive processes in people with various psychopathologies (schizophrenia, psychopathy, neurosis) show failures in the sensory systems function as compared to healthy people (Bardin et al., 1972; Korzh, 1972; Grigorieva et al, 2002). However, sensory performance in chronic somatic patients have not been properly studied. This paper intends to fill this gap.

Method

At the first stage, the "Yes - No" psychophysical method was employed to find sensory performance and decision making indices of the line length visual discrimination in a sample of 20 healthy adult people. The subjects were to compare the length of two horizontal segments of a straight line, which were demonstrated for five seconds on a PC screen. In 50% of trials the segments were equal (noise trials), in 50% segments, they differed by 10% (signal trials). The sequence of "Same" and "Different" line pairs was random. Each subject performed 50 trials (including 25 equal and 25 unequal pairs). Each session was preceded by a ten-minute training. Results were evaluated by the following indices:

- Signal detection probability: Hits (P(Y/S)) – unequal lines detection probability;
- False alarms probability: P(Y/n) – the probability of false detection of equal lines;
- Discriminability d´-index – discrimination accuracy based on P(Y/S) и P(Y/n) indices;
- Response time RT – a period between an on-screen line presentation and a subject's response;

- The decision-making criterion index (Pn), which allows to evaluate the contribution of non-sensory factors to the sensory task performance. This index was calculated using the following formula (Atkinson, 1963): Pn = P(Y/n) / (1- P(Y/S)) + P(Y/n), where 1- P(Y/S) means the probability of misses of the signal (inequality of lines).

We found no significant correlation between the performance of subjects and their age.

At the second stage, the same indices were calculated in three somatic disease groups (oncology, cerebral palsy, allergic dermatosis; 20 subjects in each group). Student's *t*-test was used to estimate the statistic reliability of performance differences between healthy participants and participants with somatic diseases. The research was carried out in in-patient hospitals in Tula, Russia.

Results and Discussion

The average comparative results of the research are shown in Table 1 and in Figures 1 - 4.

Table 1. Sensory process indices under different diseases.

Index\Group	Healthy	Oncology	Cerebral palsy	Allergic dermatosis
Signal detection probability: Hits P(Y/S)	0.79	0.84	0.61*	0.97*
False alarms probability: P(Y/n)	0.14	0.32*	0.42*	0.42*
Response time RT	1.06	1.57*	2.12*	1.04
Discriminability d'	1.88	1.46*	0.48*	2.08*
Pn	0.4	0.67*	0.52*	0.93*

Note: the asterisk indicates significant differences between patients with chronic diseases and the control group of healthy people (p<0,01).

While the signal detection probability (Hits) in the cerebral palsy group is considerably lower than among healthy people ($p<0.01$), it is significantly higher in the allergic dermatosis group compared to the healthy group ($p<0.01$). The false alarm probability is much higher in all chronic disease groups than in healthy participants ($p<0.01$). The response time is considerably greater only in oncology and cerebral palsy patients ($p<0.01$). The decision-making criterion index (Pn) is significantly higher in all chronic disease groups (because the false alarm probability is higher in them) than in healthy participants ($p<0.01$), it means that the criterion becomes more "soft" in these pathologies.

It is worth noting a slightly higher discriminability in people with allergic dermatosis. Excessive sensitivity and reactivity are typical features of such patients. At the same time the discriminability is lower in people with oncology and in those with cerebral palsy especially.

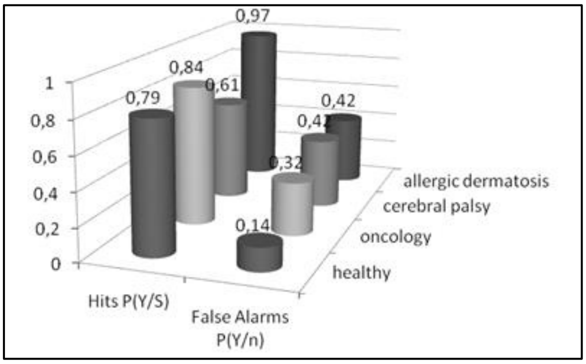


Fig. 1. Comparison of the Hits - index and the False alarms probability index under different diseases and in healthy participants.

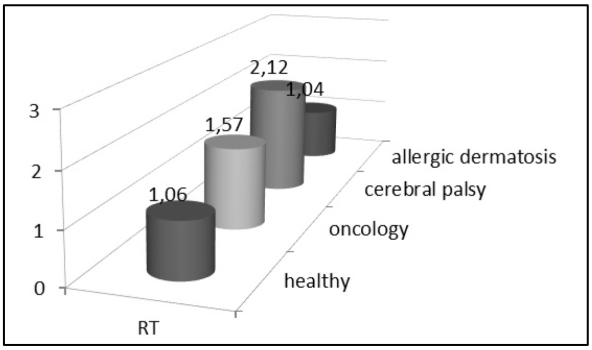


Fig. 2. Comparison of the Response time index in chronic patients and healthy participants.

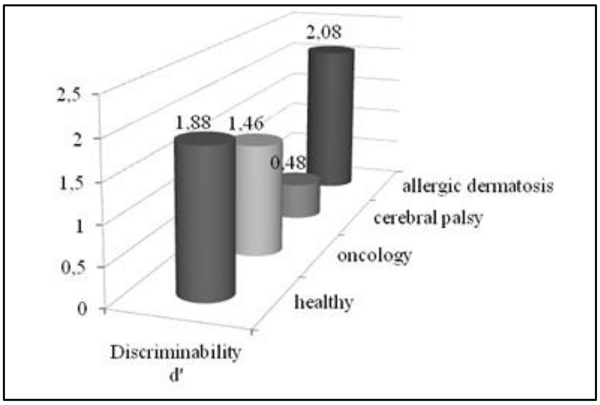


Fig. 3. Comparison of the discriminability index in chronic patients and in healthy people.

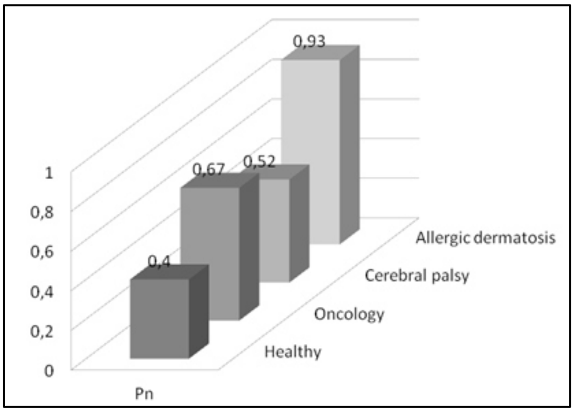


Fig. 4. Comparison of the Decision-making criterion index in chronic patients and in healthy people.

The results obtained show abnormalities in sensory systems under chronic diseases. The sensory performance in chronic disease groups changes when it comes to decision-making. The decision-making criterion is less rigorous in comparison to that of healthy people. Change in the decision-making strategy may be due to the powerful non-sensory disease factor, which influences the sensory task performance and decreases responsibility for it, as patients pay more attention to their functional state which is more important for them than to the sensory task which is not significant.

The research aimed at studying sensory performance in people with chronic somatic diseases (oncology, cerebral palsy and allergic dermatosis). The results demonstrate abnormalities in the sensory part of reception and processing of visual information and decision making systems in people with studied chronic diseases. The dynamic pattern is multidirectional: while the discriminability level in patients with oncologic diseases and especially in patients with cerebral palsy is lower than in healthy people, it is higher in patients with allergic dermatosis than in healthy people. Temporal properties of the sensory process are also abnormal: the response time is greater in the oncology group and it is even greater in the cerebral palsy group. The decision-making criterion is more “soft” in chronic somatic patients (especially in those with allergic dermatosis) than that in healthy people.

References

Atkinson R. (1963). A variable sensitivity theory of signal detection (*pp. 70, 91–106). Psychol. Rev.*

Bardin K.V., Zabrodin Y.M., Ivanitskiy A.M., Matveeva L.V. (1979). Osobennosti reguliatsii sensorno-pertseptivnogo protsessa pri nekotoryh psikhicheskih rasstroystvah. *In Psikhofizika sensornyh system. (pp. 153-189*). Moscow: Nauka. (In Russian). (Regulation of sensory and perceptive processes features in some psychopathologies. In: Psychophysics of sensory systems. *(pp. 153-189*). Moscow: Nauka).

Grigorieva L.P. (Ed). (2002). Deti s problemami v razvitii (kompleksnaya diagnostika i korrektsiya). Moscow: IKC „Akademkniga“.(In Russian). (Children having developmental problems (diagnostics and correction). Moscow: Publishing House „Akademkniga“).

Korzh N.N., Vaitkyavichus G.G. (1972) *O mekhanizme sensornyh protsessov pri obnaruzhenii signalov u bolnykh shizofreniey. In Zhurnal nevropatologii i psikhiatrii im. S.S.Korzhakova (V.72, issue 11, pp. 1689–1694).* (In Russian). (On some sensory processes mechanisms underlying signal detection in schizophrenia patients. In: Neuropathology and Psychiatry Journal after S.S. Korsakov. *V.72, issue 11, pp. 1689–169.*).

COMMAND SIGNALS AND FEEDBACK IN THE PERCEPTION OF WEIGHT AND MASS

Helen E. Ross
Psychology, University of Stirling, Stirling FK9 4LA, UK
h.e.ross@stir.ac.uk

Abstract

Several experiments show that the mass of an object can be judged by the ratio of the applied force to the resulting accelerative force. Errors can occur through failure to monitor the command signals to the muscles correctly, or incorrect evaluation of the feedback regarding the applied force, the achieved hand movements or the achieved acceleration.

The weight of an object can be judged statically through the pressure it exerts on a supported hand owing to the force of gravity. However, discrimination is much better if the observer picks up the object and lifts it up and down or moves it to and fro (Jones, 1986). In this case the observer gains information about inertial mass in addition to weight. In weightless conditions (such as spaceflight or neutral buoyancy in water) movement is essential to gain a sense of mass.

The laws of physics tell us that Force = Mass X Acceleration. Mass is therefore the ratio of force to acceleration. This formula suggests that the perception of mass is very simple: the observer applies a force to the object and then notes the achieved acceleration. Provided the observer has correct knowledge of both the applied force and the resulting accelerative force, he should correctly perceive the mass of the object. Unfortunately, the laws of physics do not map neatly onto the neurophysiology or psychophysics of weight and mass perception. Factors such as the size-weight illusion (Buckingham, 2014) and adaptation (Ross, 1981) have large effects on perception. Moreover, there is not a clear distinction in time between the command signal from the brain to the arm and hand muscles, and the feedback from the pressure receptors in the hand regarding the accelerative forces. A typical oscillatory shake is an ongoing process: with sinusoidal hand movements, there are four periods of peak acceleration per cycle (just after the start and before the end of each direction of movement). The reactive force stimulates the pressure receptors in the hand maximally at those peak times. Complex interactions occur in a fraction of a second during hand movements.

There is controversy as to whether command signals are monitored separately from feedback. The applied force could be known through feedback from muscle stretch receptors and joint receptors. The idea of ‘sensations of innervation’ originated with Helmholtz and Wundt, and led to later terminology such as ‘corollary discharge’ and ‘efference copy’ (Ross & Bischof, 1981). Some experiments are discussed below concerning the information and forces involved in hand movements and the perception of weight or mass.

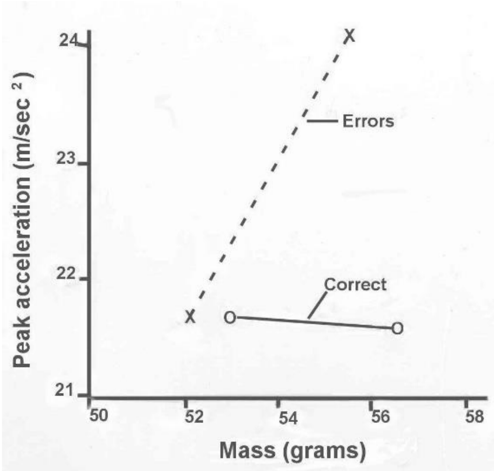


Fig. 1. Mean peak acceleration for correct and incorrect weight (mass) judgements by one astronaut postflight. Reprinted from Ross (2008).

Studies of Hand Movements

Hand movements and errors of discrimination

In a Spacelab experiment (Ross et al., 1987) astronauts were trained to produce standard hand movements, with short or long shakes, while comparing the mass of weighted balls. The short shakes (which gave higher accelerations) had intended distances of about 20 cm, and the long shakes 40 cm. Video analysis of amplitude and duration showed that the astronauts produced higher peak accelerations under zero-gravity than on the ground, and lower accelerations on the ground for a couple of days postflight. The differences were mainly due to differences in velocity rather than amplitude, although amplitudes were variable. Discrimination between the mass of the balls was also poorer inflight and postflight than preflight. A contributing factor in loss of discrimination was incomplete adaptation to changes in gravity, and this may have impaired the ability to monitor command signals correctly.

Ross (2008) found that errors in discrimination, for one observer postflight, were correlated with unusual hand movements. For correct judgements, heavier objects gave a slightly lower peak acceleration; but for incorrect judgements, lighter objects gave a much higher peak acceleration (Fig. 1). This probably implies that the observer assumed that a higher peak acceleration was due to the lighter mass of the test object, when it was actually due to a greater applied force. It is not clear whether there was only a failure to monitor command signals correctly, or whether there was also a failure in interpreting the feedback regarding the achieved movement and acceleration. The achieved hand movement could be seen visually and felt kinaesthetically, while the accelerative force could be felt by the pressure receptors in the hand.

Hand movements and subjective equality

Baud-Bovy and Socchia, (2009) measured the point of subjective equality (PSE) rather than discrimination. Participants stood in front of a haptic-robotic device, which displayed masses in a zero-gravity virtual environment, as if the masses were sliding on a frictionless table. Participants produced to-and-fro horizontal movements to assess two successively presented inertial masses. Visual and audio signals were used to impose three movements of different amplitude and/or duration for the standard and comparison stimulus. Two movements were selected to produce the same peak acceleration, and PSEs showed that the masses were judged as equal when the peak accelerations were the same, even when the movement amplitude and pace differed. However, differences in movement amplitude and frequency caused misperception of physical masses when peak accelerations differed. A lighter mass that was moved faster or over a longer distance was judged equal to a heavier mass that was moved more slowly or over a shorter distance.

Maximum force is the product of mass and peak acceleration, and this product might be a cue for mass. If two masses are perceived as equally heavy when acceleration peaks differ, the two masses should differ by the amount necessary to make the product between the acceleration peaks and the corresponding masses equal. In practice, the mass varied less than predicted if estimates were based only on maximum force.

The role of visual feedback

Research has shown that false visual feedback can affect heaviness judgements. Streit et al. (2007) found that when rods were hidden from view and wielded, and their (false) movement shown on a virtual screen, perceived heaviness varied with apparent acceleration. Rods that appeared to move faster felt lighter than those that appeared to move more slowly.

The role of muscular activity

Waddell and Amazeen (2018) argued that perceived heaviness is a function of the ratio of muscle activity to the resulting acceleration of the object. Muscle activity was measured by electromyography (EMG) from the biceps. Lifting kinematics were obtained from a motion tracking system, which recorded the positions of the observer’s elbow, forearm and wrist, and of the cylinder: velocity and acceleration were calculated from these data. The observer lifted the standard to a prescribed height, followed by a test stimulus. Speed was slow, preferred or fast. The standard was given an arbitrary weight of 100, and the observer judged the stimulus relative to that. The maximum value EMG during the upward portion of the lift was taken as the index of muscular effort. Peak angular acceleration of the lift was used as a measure of lifting acceleration. Lift time increased with mass and with slower lifting speeds. Heaviness estimates increased with mass (approximately linearly), with no effect of lift speed. Peak EMG increased with mass, and from slow to fast lifting. Peak angular acceleration decreased as mass increased, and as lift speed decreased. Thus, perceived heaviness is a function of the ratio of EMG to acceleration. Muscle activity contributed most to perceived heaviness at the preferred lift speed.

Knowledge of muscular activity

Fercho and Baugh (2016) investigated cognitive attribution of the source of an error in weight perception. Observers were told that EMG surface electrodes attached to the lifting hand were either part of a 'passive' system that recorded muscular activity, or part of an 'active' system that would apply energy to the muscle and thus influence weight perception. Both groups performed 90 training lifts of small black blocks, identical in appearance, which gradually increased in weight from 400 to 570g. This change was below the level of conscious perception. The Active group could have a plausible external source of lifting errors, while the Passive group would have to assign errors to their own lifting. This was followed by 10 lifts of a new larger red block weighing 1354 g. The lightest black block and the red block had a density of 1.81 g/ cm3. This was unusually heavy for the apparent plastic material of the black blocks, so observers would have to learn a novel material density at the start of the experiment. Each trial consisted of a single object lift from a tabletop platform with a force sensor, which allowed the measurement of vertical load force applied during lifting but before the object left the platform. Shutter glasses prevented vision between trials, so the change of blocks could not be seen.

Between the training and test trials, the electrodes were removed for the Active group, but not for the Passive group. For each lift, the initial peak rate of change of vertical load force rate and load-phase duration were calculated: these can be taken as estimates of predicted object weight. Analysis of the training lifts showed no effect of active versus passive EMG on weight predictions. Both groups scaled their lift forces to the objects, gradually adapting to the heavier weights in the series: they increased lifting forces to ensure object liftoff occurred at the same time relative to the start of the lift, regardless of the increase in object weight. For the test trials with the novel red block, the Passive group scaled their predictive load forces appropriately. However, after the EMG equipment was removed, the Active group failed to do the same: they had a smaller initial peak load force rate, and smaller load force at that time, and a longer lift duration. This may be because the Active group attributed changes in object weight to an external source (the Active EMG System), which was removed before lifting the red objects. Cognitive information may thus play a role in credit assignment, influencing weight prediction when lifting novel objects.

Conclusions

Feedback from peak acceleration plays a major role in mass perception, with lower acceleration associated with heavier objects. False feedback concerning muscular effort or the achieved movements impairs perception. The nature of command signals warrants further investigation.

References

Baud-Bovy, G. & Socchia, L. (2009). Is mass invariant? Effects of movement amplitude and duration. In: M .A. Elliott & S. Antonijević et al. (Eds.) *Fechner Day 2009* (pp. 369-374). Proceedings of the 25th Annual Meeting of the International Society for Psychophysics, Galway, Ireland.

Buckingham, G. (2014). Getting a grip on heaviness perception: a review of weight illusions and their probable causes. *Experimental Brain Research*, 232, 1623-1629.
Fercho, K. & Baugh, L. (2016). Cognitive attribution of the source of an error on object-lifting results in differences in motor generalization. *Experimental Brain Research*, 234, 2667-2676.
Jones, L.A. (1986). Perception of force and weight: theory and research. *Psychological Bulletin*, 100, 29-42.
Ross, H. E. & Bischof, K. (1981). Wund's views on sensations of innervation: a reevaluation. *Perception,* 10, 319-329.
Ross, H. E. (1981) How important are changes in body weight for mass perception? *Acta Astronautica*. 8, 1051-1058.
Ross, H. E., Schwartz, E. & Emmerson, P. (1987). The nature of sensorimotor adaptation to altered G-levels: Evidence from mass-discrimination. *Aviation, Space and Environmental Medicin*e, 58, (9, Suppl.), A148-152.
Ross, H. E. (2008). Haptic perception in space travel. In M. Grunwald (Ed.) *Human Haptic Perception: Basics and Applications* (pp. 273-280). Basel: Birkhäuser.
Streit, M., Shockley, K., Morris, A.W. & Riley, M.A. (2007). Rotational kinematics influence multimodal perception of heaviness. *Psychonomic Bulletin & Review*, 14, 363-367.
Waddell, M & Amazeen, E. L. (2018, in press) Lift speed moderates the effects of muscle activity on perceived heaviness. *Quarterly Journal of Experimental Psychology*, DOI 10.1177/1747021817739784

INFLUENCE OF A FALLING BODY MASS ON ITS PERCEIVED NATURALNESS AND PREDICTED TIME-TO-CONTACT WITH THE GROUND

Michele Vicovaro[1], Luca Battaglini[1], and Stefano Noventa[2]
[1] *Department of General Psychology, University of Padova, Italy*
[2] *Center for Educational Science and Psychology, University of Tübingen, Germany*
michele.vicovaro@unipd.it

Since gravitational force is a constant presence in our everyday life experience, one might expect that people should have an accurate knowledge of how gravity affects the motion of objects. Yet, if it were so, it wouldn't have taken several centuries from Aristotle's ideas to Galilei's experimentation to shed light on the physical phenomenon. Research in the field of intuitive physics has shown that people do not have a good intuitive understanding of the physics of gravitational motion. However, relatively little is known about people's *perceptual* knowledge of the phenomenon. Can people discriminate between physically plausible and physically implausible vertical falls when they are presented with realistic ongoing simulations of the event? Although people tend to have poor explicit knowledge of vertical fall, it seems nonetheless reasonable that they may have accurate knowledge of the phenomenon at a perceptual level, because visual perception of an ongoing physical event may allow them to draw representations of physically-based, previously experienced events.

In Experiment 1, we presented participants with virtually simulated wooden or polystyrene spheres that fell vertically to the ground from about two meters high. Participants were asked to rate the perceived naturalness of each vertical fall. Three factors were orthogonally manipulated, namely the implied mass of the falling sphere (i.e., light or heavy), the motion pattern (i.e., uniform acceleration vs. uniform velocity), and the magnitude of acceleration/velocity. Hypothetically, if participants could retrieve from memory stored representation of previously experienced vertical falls, then simulated vertical falls characterized by $\approx 1g$ acceleration should be rated as more natural than simulated vertical falls characterized by physically implausible motion patterns. The results do *not* provide support to the hypothesis that participants could draw representations of physically-based, previously experienced vertical falls. The results appear instead to reflect biased explicit knowledge of vertical fall, in that relatively low values of acceleration or velocity were judged to be natural for a simulated polystyrene sphere, whereas relatively high values of acceleration or velocity were judged to be natural for a simulated wooden sphere.

In Experiment 2, we used the same stimuli of Experiment 1, except that the sphere disappeared behind an invisible occluder at some point of its trajectory. Participants were asked to predict the time-to-contact (TTC) of the sphere with the ground, by pressing a key exactly when the impact of the lower edge of the sphere with the ground of the room would have occurred. Results showed that the estimated TTC for the simulated wooden sphere was slightly but consistently smaller than the estimated TTC for the simulated polystyrene sphere. The influence of the sphere's implied mass on participants' responses might be the manifestation of two distinct processes, namely an explicit heuristic in Experiment 1, and an implicit, automatic association between mass and falling speed in Experiment 2.

POSTER SESSION I

ACOUSTIC GNOSIS ALTERATIONS AND ITS RELATION TO COGNITION AND VISUAL SPATIAL PROCESSING PERFORMANCE IN CHILDREN

Patrícia Cunha[1], Isabella M.C. Silva[2], Elaine R. Neiva[1], and Rosana M. Tristão[3]
[1]*Psychology Institute, University of Brasília, Brasília, Brazil*
[2]*Speech Therapy and Audiology School, University of Brasília, Brasília, Brazil*
[3]*Faculty of Medicine, University of Brasília, Brasília, Brazil*

Abstract

Auditory processing disorder (APD) has been associated to different neurodevelopmental disorders potentializing general difficulties in children in a broad range of comorbidities. This study investigated the relation between auditory processing, audiologic evaluation, cognitive profile and visuomotor test performance in twenty-five children (7–14 years-old) with specific learning disorder. A complex of behavioral auditory processing evaluation was designed and included the tests dichotic digits (DD), speech in noise (SN), sound localization (LOC), and staggered spondaic words (SSW). Audiologic evaluations encompassed auditory thresholds, brainstem evoked response audiometry (BERA), P3/N2 event related potentials (ERPs). Cognitive tests included the Wechsler Intelligence Test. Bender Visual Motor Gestalt Test (BGT) was used to assess perceptual motor skills. We applied multiple linear regression to the data, and found a multitude of effects among the APE and WISC subtests, IQs, and indexes: DD and SN had significant effect in Comprehension test; LOC in Block Design; SSI in the Verbal Comprehension Index and Symbol Search test; SSW on verbal-IQ, full scale-IQ, Verbal Comprehension And Perceptual Reasoning Indexes, and Picture Completion, Information, Similarities and Picture Arrangement tests ($p<.05$). Longer BERA latencies were negatively related to WISC tests of Coding, Comprehension, Labyrinth, Picture Arrangement, and Vocabulary. Children with higher numbers of errors in the BGT had longer latencies of BERA Waves Interval I-V ($r=.693$, $t(24)=2.543$, $p=.038$). P3/N2 latencies had no effect on tests performances. APD also impacted visuomotor BGT performance, lowering the skills to perceive integration and with tendency to distort shapes. This current study reveals that acoustic gnosis is important in explaining neuropsychological outcomes, even for visual spatial processing performance, confirming the hypothesis of lower perceptual stimulus intermodal integration, visual and auditory, in children with poor auditory processing.

Difficult in reading, understanding the meaning of what is read, in spelling, with written expression and mathematical reasoning among others has been diagnosed with Specific Learning Disorder (SLD) (American Psychiatry Association, 2013). There are different factors that may interfere with the identification of SLD, such as temporal processing alterations, auditory processing disorders (APD), eye movement alterations during reading, and attention deficit and hyperactivity disorder (ADHD), among other comorbidities, which contributes to the complexity of the assessment and to the necessity of designing effective interventions. An estimated 5 to 15% of school-age children (and about 4% of adults) experience specific learning disorders. Reading disorder (dyslexia) is the most common. An estimated 70 to 80% of those with a learning disorder have a reading disorder. Around 10% of the general population is expected to be in an abnormal range for learning skills, of which 37% would be cases will present comorbidity with dyslexia and 46% comorbidity with APD (Dawes and Bishop, 2010). Though, no direct relationship has so far been found between auditory impairments and measures of cognitive skills (Rosen et al., 2010). APD also has a variety of possibilities of alterations that can be categorized in at least three different errors type as *coding, decoding,* and *binaural integration*. This study was concentrated only in children diagnosed with both APD and SLD and aimed to investigate the impact of different aspects of the auditory processing upon their cognitive and psychomotor and explored the relations among cognitive, visomotor perception, auditory processing, and short and long latencies of auditory event related potentials (ERPs).

Method

The sample consisted of 25 children with no hearing alteration, 76% males (19 children) and mean age of 10.84 years (±2.53) in a cross-sectional observational study. Written informed consent from the parents was obtained before the inclusion of their child in the study, which was performed in accordance with the Declaration of Helsinki [WHO] and approved by the Research Ethics Committee of the University of Brasília. The participants were firstly submitted to an audiologic evaluation including tone audiometry (TA), acoustic immittance and BERA measurements, long latency potentials (P3 and N2) and central auditory processing evaluation. The audiologic evaluation was conducted by tone audiometry (TA) and impedance testing with an impedance audiometer (Model AT325, Interacoustics, Assen, Denmark) (Miller, 2011). Brainstem and cortical auditory evoked potentials were recorded to evaluate auditory latency and attention sustentation skills, and were investigated by BERA and by ERPs (Model Navigator Pro, Biologic Systems, Mundelein, IL, USA). BERA yields auditory brainstem responses recorded in a series of six to seven vertex-positive waves of which I through V as well as the interpeaks I-III, III-V, and I-V were evaluated at 80 dBnHL (Eggermont et al., 2007). ERPs were recorded to detect N2 and P3 waves, via electrodes placed on the scalp (van Dinteren et al., 2014). P3 and N2 waves were elicited using an oddball paradigm in which low-probability target items (2-kHz tone bursts) are mixed with high-probability non-target (or "standard") items (1-kHz tone bursts). Secondly, the children's intelligence level and visuomotor perceptual skills were assessed. We used the Wechsler Intelligence Scale for Children (WISC), third version that is organized into different intelligence quotients (IQs) for verbal (VIQ) and performance areas (PIQ), the Full Scale IQ (FSIQ), and indexes for Verbal Comprehension (VCI), Perceptual Reasoning (PRI), Processing Speed (PSI), and Freedom from Distractibility (FDI) (Mayes and Calhoun, 2006). The Bender Visual Motor Gestalt Test (BGT) was used to assess perceptual motor skills, as it gives an indication of deficits as poor integration, perseveration, rotation >45° and distortion of the visual pattern relative to age (Reynolds, 2007).

The auditory processing evaluation (APE) was based on a validated test with norms for Brazilian children (Pereira and Schochat, 1997) using a 2-channel audiometer (Model Beta6000, Betamedical) through TDH39 phones, which includes tests as described below. The dichotic digits test (DD) to evaluate the auditory figure-ground perception ability for verbal sounds, and low performance in this test indicates an impairment of the auditory gnostic process of *decoding*. It is a list that consists of 80 digits or 20 items, each item consisting of four words representing numbers (1- 9). The test provides for the presentation of two digits in each ear simultaneously. The most frequently used test stage is binaural integration. At this stage, the individual is instructed to orally repeat the four numbers presented in both ears, regardless of the order of presentation. Correct processing of stimuli in the left ear indicates adequate inter-hemispheric communication, while altered results in both ears suggest functional alterations in the left hemisphere for speech processing. The speech in noise test (SN) assesses difficulty in understanding speech against background noise and evaluates selective and sustained attention skills, auditory closure, and low redundancy speech *decoding*. It is a repetition of 50 monosyllabic target words (i.e., two lists of 25 words each) presented monaurally against background chatter, with a signal-to-noise ratio of+10dB. The localization in five directions test (LOC) is performed to seek information about binaural *integration*, with the evaluated physiological mechanism being the discrimination of the direction of the sound source. The test consists in observing the child´s response to a rattle stimulus located above, behind, in front of, to the right, or to the left of the child's head, while they are sitting in a chair with their eyes closed. Normal sound localization ability was assumed when at least four of the five tested directions were indicated correctly. The synthetic sentence identification test (SSI) assess the auditory gnostic process of *coding* to identification of a stimuli composed by 10 sentences, presented simultaneously to a competing message composed by a children's story. The SSI test was performed with contralateral competing messages, with a main message/competing message relation of 0dB/40dB, and with ipsilateral competing messages with a main message/competing message relation of 0dB/-10dB (Bellis, 2011). The staggered spondaic words (SSW) test evaluates auditory gnostic process of *coding* involving selective attention related to the inhibition of sounds that, although present in the communication environment, are being relatively ignored. It is a dichotic test composed of 40 items, where each item consists of four words that are two pairs of paroxysmal disyllables presented either isolated or overlapping. The participant should repeat what they heard by following the order of the presentation of the words. Altered performance in this test suggests an impairment of the auditory gnostic process called coding (Katz and Tillery, 2005).

A Spearman bivariate correlation analysis was conducted to investigate interactions between sensorineural variables and WISC as well as BGT performance. A paired *t* test was run to determine whether there was a statistically significant mean difference between verbal and non-verbal intelligence test performance. Results from APE tests were compared with the WISC results, such as IQs, indexes, and tests scores, using the Mann-Whitney exact test. A multiple linear regression model was used to verify if the factors TA, APE, ERP, and sociodemographic variables would explain the WISC results. This equation did not include the BERA measurements, as these relate to lower-level brain processing and were thus analyzed as a separate measure in the Spearman bivariate correlation. All data were analyzed using SAS 9.4. A level of $p<.05$ was considered significant.

Results

The participants had an average pure-tone threshold of 10dB (±4.76), curve type A in *Immitanciometry* and no significant difference between left and right ears. The sample performance in the WISC, SON, and BGT test as well as sociodemographic profiles are described in Table 1. Mean levels of parental education (mother and father) were used, as there was no difference between parents or groups. Average values were also used for TA, latencies of BERA measurements for waves I, III, and V, and for intervals between waves and N2s/P3s when there were no differences in laterality (Table 2).

Likewise, the average for both ears in audiologic tests was used to for correlation analyses using the WISC scores. The correlation analyses between BERA measurements and WISC tests revealed a significant interaction between Wave I and Coding (r=-.653, p=.029), Waves Interval I-III and Comprehension (r=-.671, p=.017) as well as Labyrinth (r=-.698, p=.017), Waves Interval III-V and Picture Arrangement (r=-.661, p=.019) as well as PIQ (r=-.625, p=.030), and between Waves Interval I-V and Coding (r=-.653, =.029) as well as Vocabulary (r=-.619, p=.042) and PSI (r=-.654, p=.029). Children with higher numbers of errors in the BGT had longer latencies of BERA Waves Interval I-V (r=.693, $t(24) = 2.543$, p=.038). The WISC full scale IQ (115.33 ± 18.60) as opposed to the SON-R full scale IQ (98.41 ± 14.11) showed a statistically significant increase of 16.92 (95% Confidence Interval (CI), 2.045–20.778; t(1,24 =2.583, p=.020).

Subjects without alterations in the DD test had significant higher average values in the WISC Comprehension test (p=.042). The APE SN test also revealed impairments in the Comprehension test (p=.048). The LOC test revealed a negative correlation with the WISC Block Design subtest, that is, subjects with altered performance in the LOC had better outcomes in the Block Design subtest. Subjects with normal results in the SSI test had higher average results in the VCI and in the symbol search subtest (p=.036 and p=.018, respectively. Subjects with normal results in the APE SSW subtest also showed better results for eight WISC items: VIQ (p=.006), FSIQ (p=.024), VCI (p=.045), PRI (p=.015), and the subtests Picture Completion (p=.045), Information (.015), Similarities (.033) and Picture Arrangement (.033). Altered APE performance also predicted higher numbers of BGT errors (R=.64, R^2= .40, p=.049), especially on distortion, rotation, and integration.

Discussion

The results of this study show that individuals with APD and SLD have poorer scores on verbal scales and poorer factorial scores for verbal comprehension and perceptual organization, confirming our own hypothesis as well as replicating Miller's (2011) findings that cognition may be a predictor of speech recognition, communication, and listening skills. Alterations in APE performance resulted in stimulus loss and consequently in performance loss. Looking at the sensorineural dimension of the collected data, we found that long latency potentials (N2 and P3) were not altered in our sample and did not show correlations with other variables, indicating that higher-level information processing usually linked to attention, decision making, categorization, and executive cognitive control functions, is not implicated in WISC performance (see discussion section in Katz and Tillery 2005).

Table 1. Mean standard scores of sociodemographic factors, WISC cognitive profiles, and SON-R total IQ and BGT scores for type of errors.

Sample Description (N=25)		M ± SD / Nº (%)
WISC-III	*FSIQ*	*115.33 ± 18.60*
	VCI	*110.00 ± 17.65*
	FDI	*95.83 ± 14.36*
	PSI	*95.75 ± 12.46*
	PRI	*95.83 ± 14.36*
	VIQ	*110.00 ± 17.63*
	Similarities	11.75 ± 5.47
	Comprehension	14.00 ± 3.21
	Vocabulary.	12.00 ± 2.62
	Digit Span	10.25 ± 3.16
	Information	9.33 ± 3.20
	Arithmetic	8.83 ± 2.36
	Picture Arrangement	9.25 ± 4.47
	PIQ	*99.08 ± 19.00*
	Coding	8.58 ± 2.31
	Block Design	9.08 ± 3.67
	Picture Completion	11.67 ± 3.68
	Object Assembly	8.92 ± 3.17
	Labyrinth	7.21 ± 3.73
BGT	Integration	1,70 ± 1.71
	Distortion	3,48 ± 2.02
	Rotation	1,78 ± 1.90
	Perseveration	,61 ± .94

Notes: $\bar{M}$± SD= Mean and Standard Deviation; Nº(%)= Frequency (percentage); FSIQ = Full Scale Intelligence Quotient; VIQ=Verbal Intelligence Quotient; PIQ=Performance Intelligence Quotient; FDI=Freedom From Distractibility Index; PSI= Processing Speed Index; PRI= Perceptual Reasoning Index.

Table 2. Mean standard scores of audiologic evaluations and latencies for sensorineural measures of brainstem evoked response audiometry (BERA), as well as long latency potentials (N2 and P3) in the right and left hemisphere.

Audiologic Measures ($\bar{M}$ ±SD)	Latency / msec	Right ear	Left ear	p-value*
BERA	Wave I	2.08 ± .49	2.04 ± .49	.145
	Wave III	4.22 ± .48	4.28 ± .49	.277
	Wave V	6.04 ± .42	6.04 ± .38	.915
	WI I-V	3.963 ± .199	4.016 ± .282	.139
	WI III-V	1.823 ± .195	1.728 ± .234	.597
	WI I-III	2.136 ± .183	2.237 ± .185	395
ERP	N2	257.40 ± 52.50	251.51 ± 82.58	.905
	P3	323.72 ± 6.57	345.18 ± 57.21	.121

However, longer latencies of brainstem waves were related to poorer performance in verbal tests on, for example, comprehension and vocabulary, even if the brainstem conduction results were in the normal population range. Surprisingly, they were also related to tests of perceptual reasoning that measure visual and visuomotor skills such as Coding, Labyrinth, and Picture Arrangement, thereby lowering the performance IQ (PIQ) and the Processing Speed Index (PSI). These tests are associated with executive functions, and we must consider that the BERA wave results are a measure of maturity of the acoustic nerve, as well as of its integrity and its interaction with visual paths that cross in at least two structures of the brainstem. For instance, comparatively shorter latencies of wave V (possibly in the lateral lemniscus and/or the inferior colliculus) in both ears are clearly related to the ability to perceive new sounds, while longer latencies are associated with longer observation times (of the visual stimulus). This means that people with shorter latencies have a tendency towards more habituation and are better able to discriminate sound novelty (Coenraad et al., 2010). Hence longer BERA latencies, even within the normal range, should be investigated in children with SLD, as a neural immaturity marker, but also as part of follow-up programs.

Looking at general scores as well as sub-scores, we found that altered APE results were generally predicting a lower full scale IQ, specifically alterations in tests performance, such as for the SN and the SSI, predicted to generally lower comprehension abilities and SSW performance predicted both verbal and executive functions with a highly significant focus on the perceptual reasoning index. In general, normal APE evaluation seems to be linked to better results on the verbal comprehension index of WISC. Nonetheless, coping strategies might have been reflected in the relationship between poorer sound localization abilities and enhanced skills on the Block Design, a WISC subtest for spatial reasoning with concrete shapes.

In conclusion, the current study reveals that acoustic gnosis is important in explaining neuropsychological outcomes, even for visual spatial processing performance, confirming the hypothesis of perceptual stimulus loss in children with poor auditory processing (Miller, 2011). Learning to read, for instance, depends on two skills, first, being able to use spoken language and, second, proper visual object perception that are processed in a cross-modal fashion; suboptimal automatization of the integration of visual analytical skills and language processing may thus lead to learning difficulties (Lachmann et al. 2012). The WISC VIQ and the VCI showed the biggest correlation with acoustic gnosis. Although poor sound localization was correlated with better spatial processing, this might have been due to a coping process, and further studies are needed to confirm this interpretation. The findings of this study also emphasize the impact of auditory processing on verbal output and visual spatial processing. This study thus reinforces the results of earlier studies (Miller, 2011; Albuquerque et al. 2012). Moreover, it has been suggested that central auditory processing skills mainly develop until 10 or 12 years of age (Katz and Tillery, 2005), indicating that the recognition and rehabilitation of this function should happen as early as possible in the life of a child.

Acknowledgements

We are very grateful to all parents and children who took part in this study. This research was supported by a grant from FAPDF and CNPq, linked to the Ministry of Science, Technology and Innovation of Brazil (grant number193.000.326/2009) awarded to Rosana M Tristão.

References

Albuquerque, G., Maia, M., França, A., et al. (2012). Language processing in Attention Deficit Hyperactivity Disorder (ADHD). *D.E.L.T.A.,* 28(2), 245-280.

American Psychiatry Association (2013).Diagnostic and Statistical Manual of Mental disorders - DSM-5.5th.ed. Washington: American Psychiatric Association.

Bellis, T. J. (2011). *Assessment and management of central auditory processing disorders in the educational setting: From science to practice*. Plural Publishing, 2nd Ed., San Diego.

Coenraad S., et al. (2010). Fitting model of ABR age dependency in a clinical population of normal hearing children. *Eur Arch Otorhinolaryngol,* 267, 1531-1537.

Dawes, P., & Bishop, D.V.M. (2010). Psychometric profile of children with auditory processing disorder and children with dyslexia. *Arch Dis Child*, 95, 432-436.

Eggermont, J. J., Burkard, R. F., & Manuel, D. (2007). Auditory evoked potentials: basic principles and clinical application. Hagerstwon, MD: Lippincott Williams & Wilkins.

Katz J., & Tillery, K. L. (2005). Can Central Auditory Tests Resist Supramodal Influences? *Am J Audiol*, 14, 124-127.

Lachmann T., Khera G., Srinivasan N., & van Leeuwen C. (2012). Learning to read aligns visual analytical skills with grapheme-phoneme mapping: evidence from illiterates. *Front Evol Neurosci*, 4, 8.

Miller, C.A. (2011). Auditory Processing Theories of Language Disorders: Past, Present, and Future. *Lang Speech Hear Serv Sch*, 42, 309-319.

Pereira, L. D., & Schochat, E. (1997). *Central Auditory Processing: Evaluation Manual*. São Paulo, Brazil: Lovise.

Reynolds, C.R. (2007). Koppitz-2: *The Koppitz Developmental Scoring System for the Bender-Gestalt Test*, Austin, TX: Pro-Ed Inc.

Rosen, S., Cohen, M., & Vanniasegaram, I. (2010). Auditory and cognitive abilities of children suspected of auditory processing disorder (APD). *Int J Pediatr Otorhinolaryngol*, 74, 594-600.

van Dinteren, R., Arns, M., Jongsma, M. L. A., & Kessels, R. P. C. (2014). P300 Development across the Lifespan: A Systematic Review and Meta-Analysis. *PLoS ONE* 9(2), e87347.

Wechsler, D. (1991). The Wechsler intelligence scale for children - third edition. San Antonio, TX: The Psychological Corporation.

DISENTANGELING THE IMPACT OF PROSODY AND SEMANTICS IN EMOTIONAL SPEECH PERCEPTION: A COMPILATION OF EMOTIONALLY AFFECTIVE AND NEUTRAL SENTENCES IN GERMAN

Sabrina Defren[1], Patricia B. C. Wesseling[1], Shanley Allen[2], Vered Shakuf[3], Boaz M. Ben-David[3], and Thomas Lachmann[1]
[1] *Cognitive and Developmental Psychology, University of Kaiserslautern, Germany*
[2] *Psycholinguistics and Language Development, University of Kaiserslautern, Germany*
[3] *Communication, Aging and Neuropsychology Lab (CAN lab),School of Psychology, The Interdisciplinary Center (IDC), Herzliya, Israel*
sabrina.defren@sowi.uni-kl.de

The complex interplay of prosody and semantics and their role in the perception of emotion in spoken language is a topic open to discussion. Different approaches try to determine the nature of this interaction, for instance which factor dominates over the other. The present study provides the means to examine precisely this interaction in German by generating sentences with validated emotional content regarding the emotions anger, fear, happiness, sadness, or neutral. German native speakers ($N = 47$) selected the sentences based on ratings on a 6-point Likert scale in terms of emotional semantic content. The result thereof are 54 sentences, 10-11 per emotional category, balanced regarding linguistic factors (word frequency, phonological neighborhood density, and number of syllables). The sentences were recorded in different prosodies, either congruent to the semantic content or incongruent, by a professional actress. This allows an independent variation of prosody and semantics in studies with factorial design.

INVESTIGATING DIFFERENCES IN THE EFFECT OF STIMULUS ENTRAINMENT ON AUDITORY COGNITION AS A FUNCTION OF MUSICAL TRAINING.

Naomi du Bois[1], Aleksandar Aksentijevic[2], and Mark A. Elliott[1]
[1] *School of Psychology, National University of Ireland, Galway, Ireland*
[2] *Department of Psychology, University of Roehampton, UK*
mark.elliott@nuigalway.ie

Abstract

Temporal cues are fundamental to auditory perception (Shamma, Elhilali, & Micheyl, 2011). Previous visual priming research, examining oscillatory binding, has illustrated the importance of the temporal parameters of the oscillatory activity encoding stimulus features, with phase-locked activity between evoked gamma frequencies and steady-state theta rhythms priming an anticipatory response and anti-phase relations between these frequencies priming synchronous, or slightly retroactive responses (Elliott, 2014). A recent magnetoencephalographic (MEG) study (unpublished), revealed differences in evoked auditory gamma-band responses to an auditory priming stimulus with amplitude modulations in the gamma-band range (5 conditions, 31, 33, 35, 37, and 39 Hz), as a function of musical training. The deleterious effects of noise on neural synchrony have been well documented (Billings, Tremblay, Stecker, & Tolin, 2009), therefore to explore this finding further, a simultaneous auditory masking experiment was designed to eliminate the effect of entrainment. The results from a sample of musicians (n = 16) and a musically naïve control group (n = 16) will be discussed.

Scene analysis involves the segregation of the components of stimuli generated by discrete environmental events into distinct perceptual representations, or streams (Bregman, 1990; Elliott & Müller, 1998; Elliott, Shi, & Kelly, 2006). The components of these visual and auditory stimuli, such as the contour and colour of objects, and the pitch and timbre of sounds, are coded in different neural assemblies. Binding is the term given to the process of integrating these separate neural systems. This is achieved via synchronised oscillatory activity. Rhythmic vibrations in neural assemblies arising from neural feedback connections result in a synchronisation of their firing patterns (Gray & Singer, 1989; Michalareas et al., 2016; von der Malsburg, 1981). Visual and auditory priming paradigms have been designed to psychophysically examine the temporal dynamics involved in the binding process within these modalities (Aksentijevic, Barber, & Elliott, 2011; Elliott, Conci, & Müller, 2003; Elliott et al., 1998). Evidence from this line of visual binding research suggests there is an interaction in phase between an entrained oscillatory frequency and a slower endogenous cortical theta rhythm of 6.69Hz. When both oscillations are in phase alignment at the time the target is presented an anticipatory response to the primed event is preferred, that is reaction times (RTs) are faster when the target is presented just before the synchronous prime is expected in the synchronous premask condition (see Elliott, 2014, for details). However, if the interaction is out of phase, responses are faster for targets presented at the same time or slightly after the primed event, i.e., presentation times when or just after the synchronous prime within the premask is expected. Published auditory binding findings, using a similar paradigm, have found a facilitation of responses to a target sound that is inharmonically related by an octave and a minor third to the tones embedded in the prime, that is rate and time specific. RT responses are facilitated when the participant is primed with a sequence of tone bursts that are presented at 33 pips per second (pps), equivalent to a 33Hz amplitude modulation, and the target sound is presented 106ms post entrainer offset (Aksentijevic et al., 2011). This is referred to as a 'pop-out' effect, and it is suggested that it is due to a slowing of responses to harmonic targets rather than faster responses to inharmonic targets. Musicians do not demonstrate this effect (Aksentijevic et al., 2014). Similar to the priming stimuli employed in these paradigms, a 40Hz auditory steady-state response has been demonstrated to be elicited by pure tone bursts presented at a rate of 40 per second (Galambos, Makeig, & Talmachoff, 1981). Thus, the entrainer is considered to evoke an auditory gamma-band response (aGBR) that is phase locked to the entrainment frequency.

A recent unpublished study examined the difference between musicians and non-musicians (participants with no music training) at the cortical level while performing this auditory priming task, using magnetoencephalography (MEG). The current study is based on a frequency power analysis of these data. Target presentation frequency conditions were 31, 33, 35, 37, and 39pps. Non-musicians exhibited higher amplitude gamma activity during the 31pps target presentation frequency condition, specifically 31, 36, and 41Hz activity. However, 33 and ~38Hz were reduced in amplitude across all conditions (figure 1). Musicians demonstrate gamma peaks in the 35 and 39pps conditions (not observed in the non-musician group), yet 35 and 39Hz is reduced in amplitude in all conditions. During the 31, 33, and 35pps presentation conditions, 31Hz is dominant. However, during the 33 and 39pps conditions the matching frequencies are attenuated (figure 2). When correspondence is drawn between the RT data and the frequency power analysis, it is suggested that for non-musicians the high gamma activity during the 31pps condition is coupled with the largest difference between harmonic target present (HTP) and inharmonic present (ITP) responses, while the difference was also large for responses in the 33, and 39pps conditions when gamma activity was reduced. The latter observation is possibly due to a phase interaction with a theta rhythm (~6Hz), while the 31pps difference also implies an interaction with theta (~5Hz), given the frequency difference between the entrained 31Hz gamma frequency and the observed high amplitude activity for 31, 36, and 41Hz during this entrainment condition.

Musicians demonstrated the least difference in responses to both harmonics at 33 and 39pps presentation, and in both cases, these frequencies themselves were attenuated. Again, it is implied that this is due to an interaction in phase with a theta rhythm (~6Hz). Arguably for both groups, a response to harmonically related stimuli depends on an interaction in phase between entrainment frequencies and a theta rhythm of approximately 5-6Hz. However, the effect is very different depending on music ability, as reduced gamma, presumably due to an interaction in phase with a theta rhythm, appears to disrupt the harmonic interpretation for non-musicians, but not musicians.

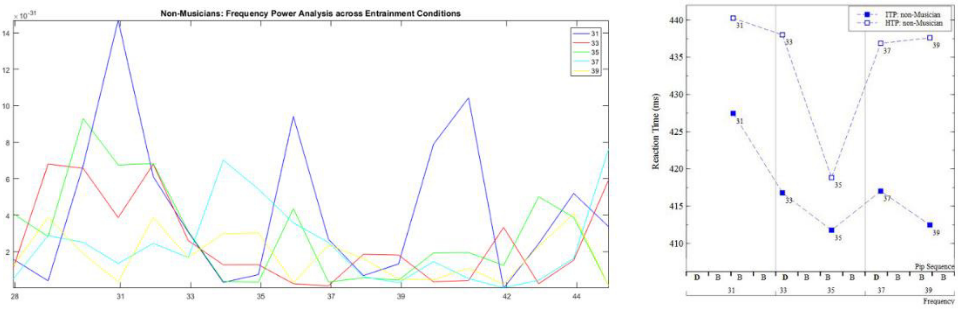


Figure 1. Unpublished MEG data. **Left**: Graphed frequency power analysis on the non-musician data during the 100ms interval between entrainer offset and target onset, bandpass filtered between 28 and 45Hz. Each line represents a different entrainment condition, with amplitude on the *y*-axis and frequency on the *x*-axis. **Right**: ITP and HTP responses at each rate of entrainment for the non-musician group, with the presentation time in relation to the oscillatory trace of the four-pip entrainer sequence on the *x* axis, and RT on the *y* axis.

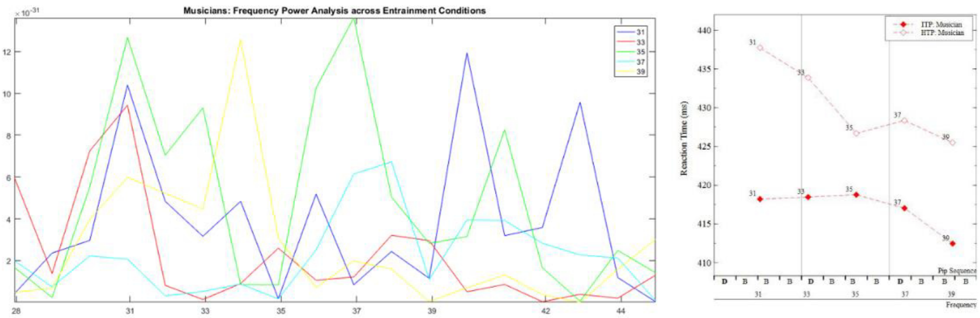


Figure 2. Unpublished MEG data for musicians graphed as described in figure1.

According to Mussachia et al. (2007), musicians demonstrate practice-related changes as training leads to reciprocal efferent and afferent plasticity which strengthens subcortical and cortical centres (Musacchia, Sams, Skoe, Kraus, & Merzenich, 2007). Thus, it is suggested that musicians depend less on early auditory binding processes and by extension, the effect of priming is possibly reduced in this population. Therefore, it is considered that by disrupting the entrainer with a noise masker, the difference in RTs to harmonically related stimuli, as a function of music training, can be compared when the auditory system is primed with gamma-band frequencies and when it is not primed. It is hypothesised that musician’s performances do not benefit from entrainment. Therefore, the musician’s responses will not differ significantly across masked and unmasked conditions, while non-musicians results will.

Method

Participants

Thirty-two participants recruited in the Galway area took part in this study. The sample consisted of two groups: Musicians with grade 6+ or an international equivalent ($n = 16$), and non-musicians with no music experience ($n = 16$). All participants reported normal hearing and provided informed consent.

Design

A mixed design was used with between factor, group (musicians and non-musicians), and within factors rate (5 levels; 31, 33, 35, 37, and 39Hz), mask (mask and no mask), and target (asymmetrical: Target absent (TA) on 50%, HTP on 25%, ITP on 25%, of trials). It can be challenging to discriminate between HTP and ITP targets and therefore a TA condition is included so that participants can make a response more easily, responding TP for both HTP and ITP targets. TA responses are not of interest and are removed from analyses. Mask and unmasked sessions took place on separate days and were counterbalanced across participants. Experimental conditions were randomised within and between blocks. The dependent variable was RT.

Paradigm

The auditory priming paradigm used was designed by Aksentijevic, Barber, and Elliott (2011). Each trial consists of a sequence of two sound stimuli (pip-trains). The first is the entrainer, which carries a repeated sequence of a four-pip prime; one 1,000Hz deviant pip followed by three 500Hz baseline pips. The second is the target stimulus which contains two different conditions – absent or present. The target absent (TA) pip-train consists of only 1,000Hz pips while the target present pip-train carries alternating pips which were either harmonically related to the entrainer (HTP, 1000Hz and 2000 Hz) or inharmonically related (ITP, 1000Hz and 2400 Hz). Entrainer duration and inter-stimulus interval (ISI) were consistent with Aksentijevic et al., (2011). Readers should refer to Aksentijevic et al. (2011) for details. Sound stimuli were presented diotically, and stimulus intensity was kept at an average 40dB sound pressure level (SPL; A weighting). 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. During the masked session, the entrainer on each trial was simultaneously masked with broadband noise, with an increased stimulus intensity of 15dB compared to the entrainer and target stimuli, i.e. measuring an average 65dB sound pressure level (SPL; A weighting). Calculated in Matlab, the signal to noise ratio (SNR) for the combined entrainer and broadband noise was 2.5dB, which is below the recommended SNR for signal thresholds (between 5 and 15dB; Yost, 2001).

Results and Discussion

To test the effect of priming on RT responses to a harmonic relation as a function of music training, a mixed ANOVA was conducted for the between and within factors on RT responses. While the effect of group was not significant ($F_{(1,30)} = .6$, $p = .44$,

η_p^2 = .02), significant main effects were found for harmony, and rate; $F_{(1,30)}$ = 35.1, p < .001, η_p^2 = .54, and $F_{(4, 120)}$ = 3.38, p = .012, η_p^2 = .1 respectively. Also, there was a significant interaction effect between group and harmony ($F_{(1,30)}$ = 6.8, p = .014, η_p^2 = .54), and between group, mask, harmony and rate ($F_{(4, 120)}$ = 2.93, p = .024, .001, η_p^2 = .1). The main effect for harmony was due to overall faster RTs to ITP targets (*MD* = 14.56ms, *SE* = 2.46ms, p < .001), while the effect for rate was due to RTs during the 35pps condition that were significantly faster compared to RTs during the 31 and 33pps conditions (*MD* compared to 31pps RTs = 5.15ms, *SE* = 1.59ms, p = .029, and *MD* compared to 33pps RTs = 5.69ms, *SE* = 1.59ms, p = .046). Bonferroni adjusted pairwise comparisons revealed that the four-way interaction was due to a significant difference in HTP RT responses that was faster for non-musicians compared to musicians in the masked condition, at a presentation rate of 31pps (*MD* = 52.27ms, *SE* = 24.88ms, p < .044).

To examine the effect of entrainment versus no entrainment more closely, the alpha level was adjusted to .025, and a further ANOVA was carried out on the data for each mask condition separately. When the entrainer was not masked, i.e. under normal entrainment conditions, there was a significant effect of harmony, rate ($F_{(1,30)}$ = 23.8, p < .001, η_p^2 = .44, and $F_{(4, 120)}$ = 3.4, p = .011, η_p^2 = .1, respectively), and a significant three-way interaction between group, harmony and rate ($F_{(4, 120)}$ = 4.67, p = .002, η_p^2 = .14). The effect of harmony was due to faster ITP RTs overall (*MD* = 16ms, *SE* = 3.28ms, p < .001), while the effect of rate was due to faster overall RTs as rate increased. While ITP RTs were faster at all entrainment rates for the musician group, the interaction effect was due to ITP RTs that were significantly faster compared to HTP RTs for the non-musician group at 33pps entrainment only (*MD* = 18.51ms, *SE* = 5.41ms, p = .002) – evidence of a pop-out effect which is consistent with the previous literature. The difference for this group at 39pps approaches significance (p = .027), which is consistent with the RT effects for the unpublished MEG results. The only difference between the findings of this experiment and previous results using this paradigm was that HTP responses at an entrainment rate of 31pps were faster than ITP RTs at this rate for the non-musicians (ITP responses in all other studies have been consistently faster than HTP responses), and HTP RTs, perhaps as a consequence, were significantly slower in the musician group compared to the non-musician group at this rate of entrainment. Interestingly, as observed in the frequency power analysis on the MEG data, 31 Hz entrainment was very clear for the non-musician group, and 31Hz, in general, was high in amplitude across entrainment conditions 31, 33, and 35pps for the musician group. The implication is that entraining a 31Hz auditory gamma-band response (aGBR) increases the amplitude of this information carrier, perhaps enhancing non-musician's responses.

The results were quite different when the entrainer was masked by noise. The only main effect is for harmony ($F_{(1, 30)}$ = 26.63, p < .001, η_p^2 = .47), due to faster ITP RTs for both groups (musicians; *MD* = 16.1ms, *SE* = 3.59ms, p < .001, and non-musicians; *MD* = 9.8ms, *SE* = 359ms, p = .01). Bonferroni adjusted pairwise comparisons reveal the HTP-ITP difference is significant at all levels of rate for the musician group, while the difference only remains significant for the non-musician group at entrainment rates 31, and 37pps. It can also be observed that RTs do not decrease with increasing rate for either group and while not significant, ITP responses for musicians do not differ overall between mask and no-mask conditions, while HTP responses are faster overall when the entrainer is masked.

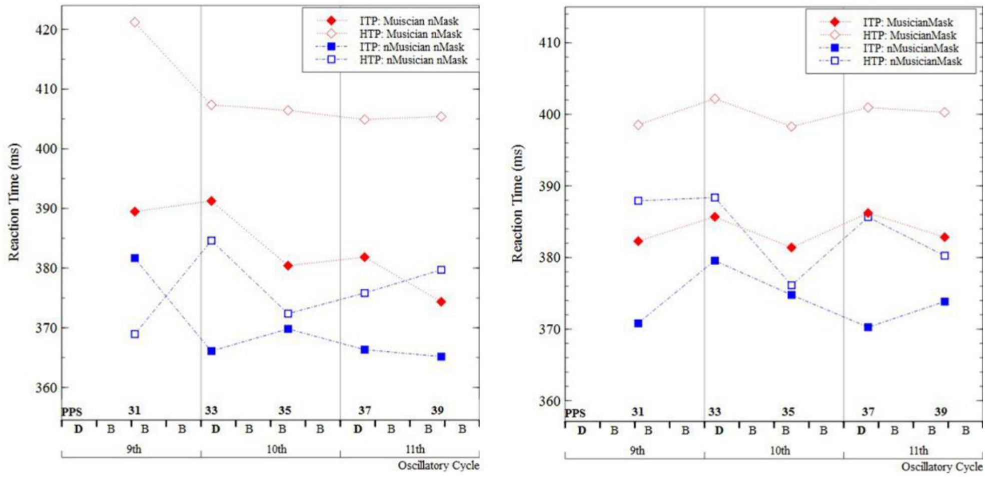


Figure 3. Left: Illustration of ITP and HTP responses for musicians and non-musicians under normal entrainment conditions, at each rate of entrainment, mapped over the oscillatory trace of the four-pip entrainer sequence (on the *x* axis), with RT on the *y* axis. Right: Illustrated as on the left, only when the entrainer is masked with broadband noise.

The results support the hypothesis that entrainment does not benefit musicians, however, it may slow responses to harmonically consonant targets, while entrainment does elicit differential effects when participants do not have any music training, entrainment at a specific rate of 33pps (and to a lesser extent at 39pps) facilitates a response to inharmonically related tones, while responses at 31pps entrainment are enhanced. The findings also support the interaction between 33 and 39Hz and a slower ~6Hz rhythm implied by the frequency power analysis conducted on the MEG data for non-musicians. Interestingly when the system is not primed with these gamma frequencies, the non-musician group demonstrate a shift in the ~6Hz interaction from 33 and 39 Hz to 31 and 37Hz.

Music training induces practice-related changes in cortical processing, presumed to underlie pitch perception. It has been suggested that this is due to a combination of top down and bottom up influence. The former is the result of modified neural architecture required for performance, due to music training, beginning with higher processing areas and gradually enhancing lower sensory areas. The latter is a consequence of Hebbian learning, which arises from the simultaneous activation of pre and post-synaptic auditory brainstem neurons, strengthening the efficacy of the brainstem responses encoding sound (Musacchia et al., 2007). The findings of this experiment support this hypothesis. Effects of entrainment occur at an early stage of processing, which benefits a response to a harmonic relation in the non-musically trained brain, under specific temporal parameters, presumably as they rely more on early auditory processing for sound perception. The effect is greatly reduced for musicians, presumably due to top-down influence resulting from music training.

Acknowledgements

This research was sponsored by the Irish Research Council (IRC).

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*.

Aksentijevic, A., Smith, A., & Elliott, M. A. (2014). Rate-specific Entrainment of Harmonic Pitch. *Music Perception: An Interdisciplinary Journal*.

Billings, C. J., Tremblay, K. L., Stecker, G. C., & Tolin, W. M. (2009). Human evoked cortical activity to signal-to-noise ratio and absolute signal level. *Hearing research*, *254*(1-2), 15-24.

Bregman, A. S. (1990). *Auditory scene analysis : the perceptual organization of sound*. MIT Press.

Elliott, M. A. (2014). Atemporal equilibria: Pro- and retroactive coding in the dynamics of cognitive microstructures. *Frontiers in Psychology*, *5*, 1–13.

Elliott, M. A., Conci, M., & Müller, H. J. (2003). Prefrontal cortex and the generation of oscillatory visual persistence. *Behavioral and Brain Sciences*, *26*(6), 733-734.

Elliott, M. A., Müller, H. J., & Miiller, H. J. (1998). Synchronous Information Presented in 40-Hz Flicker Enhances Visual Feature Binding. *Source: Psychological Science*, *9*(4).

Elliott, M., Elliott, M. A., Shi, Z., & Kelly, S. D. (2006). Title A moment to reflect upon perceptual synchrony A Moment to Reflect upon Perceptual Synchrony. *Journal of Cognitive Neuroscience*, *18*(10), 1663–1665.

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*.

Gray, C. M., & Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Sciences*.

Michalareas, G., Vezoli, J., Van Pelt, S., Schoffelen, J. M., Kennedy, H., & Fries, P. (2016). Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron*, *89*(2), 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.

Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences*, *34*(3), 114–123.

Von der Malsburg, C. (1981). The correlation theory of brain function Internal Report, Max-Planck Institute for Biophysical Chemistry Goettingen.Yost, W. (2001). Fundamentals of hearing: an introduction.

Yost, W. A. (2001). *Fundamentals of hearing: an introduction*. Academic Press.

SELF-CONTROL DETERMINES OBSERVER'S STRATEGIES IN LOUDNESS DISCRIMINATION

Svetlana A. Emelianova and Aleksey N. Gusev
Lomonosov Moscow State University, Russia
oly_e@mail.ru; angusev@mail.ru

Abstract

The central contradiction in the threshold task performance is the conflict between the necessity to discriminate signals effectively and observer`s available resources. This contradiction can be resolved in the form of additional efforts directed to compensate for resource deficiency or, in contrary, in moving away from the task and an intention to reduce resource costs. The purpose of our research is to clarify the role of the mechanisms of personal self-control determining the observer`s strategies that are used for the loudness discrimination task. The psychophysical research on loudness discrimination of tonal signals (2AFC method) has been carried out, N=106. The influence of self-regulation on RT and sensory sensitivity index A' was found out. The qualitative analysis of individual ways or manners of signals discrimination was carried out. The relationship between loudness discrimination effectiveness and self-regulation processes characteristics mediating the sensory task decision were revealed. The results are discussed in the context of understanding the individual characteristics of self-regulation as a way to overcome the situation of sensory uncertainty.

Keywords: psychophysics, individual differences, loudness discrimination task, self-regulation, additional sensory cues.

The terms for solving a threshold sensory task define its specific features comparing with other person's cognitive tasks: 1) sensory information deficit; 2) unpredictable and random for an observer nature of stimulus presentation; 3) increased information loading upon an observer. Such terms cause high information uncertainty for an observer and a necessity for a high concentration of attention to the presented stimulus material as well as some efforts to support arbitrary attention in time. Solving of this sensory problem is subjected to information loading on a person that is confirmed by studies on mental stress assessment caused by different experimental tasks [2].

Efficiency of solving the threshold sensory tasks is conditioned by the influence of situational factors – the terms for solving a sensory task – and individual features of observers. Thus, we have clearly established the influence of some situational and individual factors on changing of the level of observers' activity: time of the day, duration of experience, sleep deprivation, extraversion and introversion, etc. [2]. In this work we are considering the influence of the following situational and individual factors on the efficiency of solving the threshold and near-threshold sensory discrimination tasks: difficulty level of a sensory task, features of self-regulation of a tested subject (observer).

Observer`s performance in the near-threshold area takes place under the conditions of sensory information deficit and high rate of stimulus presentation.

Therefore, the central contradiction in the performance of a threshold task is an intrapsychic conflict between the necessity to reach the goals provided by the instruction, for example, to effectively discriminate between signals, and available resources of a system of information processing. This contradiction may be clearly revealed in the form of additional efforts directed to compensate the resource deficit and to overcome stimulus uncertainty or, in contrary, in a desire to reduce resource costs and in moving away from activity. A few studies showed a role of the process of motivational and volitional regulation when performing psychophysical tasks on discrimination and detection of sensory signals [3; 6, etc.]. We assumed that in order to explain the mechanisms of solving of the conflict described above it is useful to apply theoretical frame of a metacognitive model to control an action by J. Kuhl [7]. According to Kuhl's conception and ideas, a process to control an action – in this case a sensory action – is mediated by a strategy actively realized by the subject that is expressed in orientation to the action taken or to one's own state.

Methods and procedure of the research

Observers

The research comprises 106 subjects (average age is 31 years old), 88 women and 18 men.

Stimulation

Tonal signals lasting 200 msec with a frequency of 1000 Hz. An interval between trials was 3 sec, an interval between stimuli - 500 msec. The difference between stimuli in different series were equal to 1, 2 (basic) or 4 dB (training).

Equipment and software

Registration of the responses and stimulus presentation was done with the help of personal computers with standard audio cards and stereo earphones AKG (K-44). Motor responses were recorded through special consoles that could ensure accurate registration and avoid measurement errors of reaction time (RT). Sound stimuli were synthesized in program "Sound Forge 4.5".

Procedure

A psychophysical research of loudness discrimination of tonal signals (method 2AFC) was carried out. The observer was asked to listen to two sound signals and decide which of them – the first one or the second one – is louder. Within two days a participant of the experiment took part in two tests corresponding to a simple (2 dB) and more complicated (1 dB) signal discrimination task. Each separate test included training and introductory series (20-60 trial with a difference of 4 dB) and main series that consisted of four blocks each having 100 trials. In case the observer carried out the training series without errors he moved on to the main series.

Upon completing each of the blocks of trial presentation the observer was shown the outcomes of his work – a screen displayed an estimated probability of correct responses, percentage of correct responses and false alerts. There was a break after. During that break, the observer shared with the tester his subjective impressions that he experienced during implementation of the task. Reply protocol of the observer was recorded via voice recorder and was decoded after. If the observer detected the characteristics, different from loudness, in sound of the stimuli presented to him during the test, the observer filled out a standard self-control questionnaire. Before the research the observers filled out a questionnaire “Control over an action” (HAKEMP-90) adapted by S.A. Shapkin [5].

The following sensory task performance indices were calculated for each series: 1) sensory sensitivity index A′, 2) RT. The data obtained were processed with the help of one-factor analysis of variance (ANOVA) in IBM SPSS statistical package for Windows 17.0. Independent variables (factors) were 3 scales of the factor “Control over an action”: “Control over an action when planning”, “Control over an action in case of failure”, “Control over an action in case of action realization”. Each subfactor was presented by two levels – «state orientation» (OS) and «action orientation» (OA). In order to determine the levels of the factors, the values obtained through the scales of HAKEMP-90 questionnaire (“Control over an action”) were split at the midpoint, i.e. the groups of OA-observers and OS-observers were defined for each scale.

Results

The inter-group comparison between average values of performance indicators of threshold and near-threshold sensory tasks showed that among Action Oriented observers (in comparison with State Oriented) more stable motor responses are more prevalent, i.e. lower values of mean-square deviation of RT for all types of responses when implementing a “simple” task (difference between stimuli - 2 dB). We also established that among State Oriented observers average RT during the test is higher in general than among Action Oriented observers. It means that in general they spend more time to discriminate between loudness of signals. It is shown that when solving a more complicated threshold sensory task (difference between stimuli - 1 dB), OS-observers demonstrate a higher level of differential hearing than OA-observers ($F(1.78)=7.341$; $p=0.008$). Moreover, in a “simple” task the main effect of “Control over an action” factor was not significant for sensory sensitivity indicator – sensory sensitivity indicators between two groups were the same.

Analysis of a large amount of self-report data made it possible to separate and present general specific features of individual ways of work of OA- and OS-observers. We found out that in general OS-observers mentioned their emotions more often, they described the feelings they experienced when they were having difficulties or were successfully implementing certain blocks of trials, they alluded to the features of their functional states. Earlier the works by K.V. Bardin et al. showed that auditory sensory task is resolved based on additional sensory cues occurring during the audition of sound stimuli [1: 104-108]. It is reflected in catching of “additional sensory cues” (ASC) – modal and nonspecific characteristics which represent sensory qualities of not only auditory, but other modality acoustic features – sensory qualities of auditory modality [1: 117-123]. We found out that most of ASC determined by OS-observers represented complex kinetic space images, visual color experiences, but not every ASC determined could be used, table 1. In group of modal specific, acoustic cuses the chipboards connected with intonational and voice-frequency characteristics prevailed. OA-observers, on the contrary, were focused on the task implementation and followed the instruction literally. In comparison with OS-observers OA-observers used small sets of ASC or did not use them at all applying the methods of work that excluded ASC involving, partially or completely. While sharing their impressions during the break, OS-observers told about difficulties to initiate the task, made a lot of explanations discussing their work during the test, gave examples from their everyday life, paid close attention to the reasons why the task is implemented successfully and to their failures. Unlike them OA-observers did not find thoughts and feelings in their subjective emotions that could affect or prevent them from realization of the activity initiated. In our opinion, mentioned specific features of the groups of observers being compared naturally explain the differences between the indicators of response time.

However, State Oriented observers showed higher efficiency in comparison with OA-observers when solving a more "complicated" threshold task. This advantage was reflected in a sensory component of task resolution. As we assume, a higher level of sensory sensitivity in a group of OS-observers may be considered as a proof that a greater amount of cognitive resources were involved to resolve a threshold task.

Table 1. Additional sensory cues, nonspecific characteristics. Data of self-reports.

Additional sensory cues, nonspecific characteristics	Observers (%)	
	OS-	OA-
Images of concrete objects	87.2	72.5
Graphic schemes	89.0	49.0
Size	92.7	80.3
Verbal designation of signals, naming	85.4	58.8
Vision length	30.9	35.2
Arrangement in surrounding space	43.6	41.1
Localization of a sound in head space	50.9	25.4
Direction of the movement of a sound	69.0	9.8
Color feelings	56.3	15.6
Propriotseptivny feelings	7.2	3.9
Kinestetichesky and tactile feelings	16.3	13.7
Sounding duration	60.0	29.4
Brightness, intensity of a vision	90.9	62.7
Capacity	23.6	13.7
Density	7.27	9.8
Speed of increase in loudness	38.1	50.9
Accent, emphasis on louder signal	69.0	52.9

Discussion

The data presented in this research are consistent with the outcomes obtained during studying of psychological mechanisms of resolving tasks to detect/discriminate between visual and auditory signals and it showed that variation of the type of stimulus uncertainty leads to transformation of a functional system of signal detection. It its turn it is reflected in changing of operational composition of observer's activity [4: 154-155]. In general, the data obtained comply with the Theoretical model of multidimensional sensory space by Y.M. Zaborodin as well as with the Model of compensatory discrimination/detection mechanism offered of K.V. Bardin [1: 108-114]. It suggests that when there is a great difference between stimuli in the process of taking decisions in relatively simple near-threshold sensory tasks one basic sensory axis takes part where all sensory impressions are distributed according to “loudness” parameter. To resolve a more complicated sensory task effectively when the difference between stimuli is very small one single cue – loudness – is not enough. Then with the help of determination and application of ASC of sensory images by observers new axes of sensory space begin to be formed.

Therefore, observer’s feeling of high informational uncertainty conditioned by low stimulus intensity or their minor differences between them makes the observer to choose and determine a resolution strategy. We showed that when solving threshold or near-threshold sensory tasks sensory information is not the only factor and in case of the deficit thereof not the main one that condition the outcome. Analysis of psychological mediation of sensory task resolution shows they are conditioned by observer’s activity. What is especially interesting is an answer to the question about how this activity is expressed. Control over decisions and actions in an uncertain situation is associated with processes of setting goals and motives by an observer as well as with observer’s estimation of his own efficiency. When the level of self-regulation is low the effect of extraneous factors in a situation of uncertainty is higher that leads to unstable response time. If the level of self-regulation is high, the observers are more “focused” on the task and on its instruction, but they pay less attention to their own experiences. Experimental and theoretical approach to an observer as a subject of psychophysical measurement makes it possible to describe and study observers’ methods to obtain sensory information in a situation of uncertainty.

Acknowledgements

The reported study was funded by RFBR according to the research project №18-013-01051\18.

References

Bardin K.V., Indlin Yu.A. Beginnings of subject psychophysics. Moskva: Izdatelstvo «IP RAN» (in Russian)

Gusev, A.N. (2004). Psychophysics of sensory tasks: the analysis of behavior of the person in an uncertainty situation. Moskva: MGU, UMK «Psihologija» (in Russian).

Skotnikova I.G. (2008) Problem of subject psychophysis. Moskva: Izdatelstvo «IP RAN» (in Russian).

Utochkin, I.S., Gusev, A.N. (2007). Formation of functional body at detection of a threshold signal in the conditions of spatial uncertainty. In: Psihofizika segodnja. Moskva: Izdatelstvo «IP RAN». P. 309—319. (in Russian).

Shapkin, S.A. (1997). Experimental studying of strong-willed processes. Moskva: Smysl. (in Russian).

Gusev A.N., Shapkin S.A. (2001) Individual differences in auditory signal detection task: subject-oriented study // Fechner Day 2001 / Ed. by E. Sammerfeld, R. Kompass, T. Lachmann. Lengerich: Pabst Science Publishers. P. 397-402.

Kuhl, J. (1992). A theory of self-regulation: action versus state orientation, selfdiscrimination, and some applications. Applied Psychology: An International Review, 41, 95—173.

PERCEPTUAL CONTROL OF THE IRRELEVANT SPEECH EFFECT AS A RESULT OF TRAINING ON A DICHOTIC LISTENING TASK

Florian Kattner, Stefan Coels, and Wolfgang Ellermeier
Technische Universität Darmstadt
kattner@psychologie.tu-darmstadt.de, ellermeier@psychologie.tu-darmstadt.de

Abstract

Task-irrelevant speech is known to interfere with the concurrent memorization of verbal information in serial order (irrelevant speech effect). It has been argued that such auditory distraction may result either from perceptual interference produced by automatically processed fluctuations in the irrelevant sound or from attentional capture due to salient changes in auditory stimuli. There is evidence that distraction produced by attentional capture can be reduced via cognitive control, whereas perceptual interference cannot. Here we investigate whether the perceptual interference produced by irrelevant speech can be reduced by enhancing auditory filtering skills through a dichotic listening training task. Specifically, across five training sessions, participants learned to process only letters spoken by a target voice in the shadowed ear while ignoring the remaining 75% of letters presented either by the non-target voice or in the unattended ear. The average memory span for target letters increased with training, and the disrupting effects of irrelevant speech on serial recall of visually or acoustically presented items decreased from pre-test to post-test. In contrast, the irrelevant speech effect was not reduced in an active control group that was trained on an unrelated auditory interval discrimination task. The results suggest that practicing auditory filtering is suited to counteract the perceptual interference produced by irrelevant speech.

Verbal short-term memory is known to be sensitive to the presence of background sound. Task-irrelevant speech was shown to produce particularly strong effects of auditory distraction resulting in impaired serial recall of verbal information (irrelevant speech effect; Salamé & Baddeley, 1982). It has been argued that auditory distraction may be based on either (a) perceptual interference between automatically processed irrelevant sound and the process of serial rehearsal or (b) attentional capture due to an unexpected, deviating auditory stimulus (Hughes, Vachon, & Jones, 2005, 2007). In line with this dissociation, there is recent evidence suggesting that enhanced cognitive control can eliminate the attentional capture effect resulting from a sudden acoustical deviation in the background speech (e.g. by increasing perceptual load or providing foreknowledge), whereas the perceptual interference between continuously fluctuating irrelevant speech and serial recall was not sensitive to these manipulations (Hughes, Hurlstone, Marsh, Vachon, & Jones, 2013). However, it has also been found that background speech does not produce any interference with serial recall in blind listeners, suggesting that enhanced auditory attention and perceptual control may shield against the perceptual interference produced by task-irrelevant sound (Kattner & Ellermeier, 2014). Based on these findings, it may be assumed that the perceptual interference between task-irrelevant speech and serial rehearsal can be eliminated by a training of auditory perceptual control capabilities enabling the listeners to filter irrelevant auditory information. To test this assumption, participants in the present study were trained in a dichotic listening task requiring participants to selectively attend to and memorize relevant spoken items while filtering out the irrelevant spoken information (depending on ear location and voice). The degree of interference between irrelevant speech effect and performance on a serial recall task was assessed before and after training. If the perceptual interference between irrelevant speech and serial rehearsal is sensitive to perceptual control, then the irrelevant speech effect should be reduced after this dichotic listening training, as compared to an active control group that was trained in an unrelated auditory interval discrimination task.

Method

Participants

A total of 45 participants (36 women, 9 men) were recruited at the campus of Technische Universität Darmstadt. Ages ranged between 18 and 31 years (M = 21.9; SD = 2.7). Participants were randomly assigned to either the Training group (n = 23) or the Control group (n = 22). All participants provided written informed consent and were compensated with course credit. The study was approved by the ethical review committee of Technische Universität Darmstadt on 21th April 2017 (EK 11/2017).

Apparatus

The experimental routines were programmed in Matlab (Mathworks, Natick, MA, USA) utilizing the Psychophysics Toolbox 3 extensions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). Sounds were D/A converted (16 bit, 44.1 kHz) by an external sound card (RME multiface II), passed through a headphone amplifier (Behringer HA 800 Powerplay PRO-8), and presented via headphones (Beyerdynamics DT-990, 250 Ohm) in a single-walled sound-attenuated listening booth (International Acoustics Company). Visual information was displayed on 22” LCD monitor (iiyama ProLite E2207WS) in the listening booth.

Procedure

The experiment comprised seven individual sessions for each participant. The first and last sessions served as the pre- and post-tests for the irrelevant speech effect, and the sessions 2-6 were the training sessions. Two subsequent sessions were separated by at least one day.

In the pre- and post-test sessions, participants were asked on each trial to memorize a series of eight randomly drawn digits between 1 and 9 (without replacement) that were presented either visually numbers on the screen or spoken via headphones (ca. 72 dB), on half of the trials each. The digits were presented 1/s (without inter-stimulus interval for the visual stimuli), and there was a blank retention interval of 6 s after the final digit and before participants were asked to click the series into a numeric pad that was shown on the screen. During both the presentation of the digits and the retention interval, irrelevant sound (i.e., continuous speech from a male or female speaker, or Gaussian noise in one third of the trials each) was presented over headphones (ca. 60 dB). With 10 repetitions for each condition (irrelevant sound × modality of digits), the entire session consisted of 60 trials plus two practice trials at the beginning of the session. The order of trials was randomized for each participant.

In the five training sessions, participants in the Training group were trained with a dichotic listening in which two independent lists of spoken letters (randomly drawn from the 15 letters B, F, G, H, K, L, M, N, P, Q, R, S, T, W, and X) were presented via headphones to the left and right ear. Half of the letters of each list were spoken by a male voice and the other half of letters was spoken by a female voice. Participants were instructed to attend only to the letters spoken by the male or female voice throughout the entire training session with the target voice varying between successive training sessions. On each trial, participants were asked to attend to either the letters spoken by the target voice in the left or the right ear (target letters). After presentation of the first lists, there was a short delay (2 s) before a second pair of lists of letters was presented to the left and right ear. The second target letters were either identical to the first target letters or not (i.e., one letter was replaced by a different letter). The participants' task was to indicate whether the two target lists were identical or not by hitting the "down" or "up" arrow keys on the keyboard, respectively. The length of the target lists (as defined by voice and ear) varied after each 12-trial block as a function of the participants' performance: The list length was increased by 1 letter when participants were 90% correct ($\geq$ 11 correct trials), and it was decreased by 1 letter when participants did not reach more than 70% correct ($\leq$ 8 correct trials). In the first training session, participants started with 4-letter lists, and participants could choose their starting list length in the following training sessions. The minimum list length was 2 letters. After each 12-trial block, participants were shown the list length of the next block, and they could choose to take a short break before continuing. Participants were trained for exactly 60 min in each session (not including break durations).

The Control group was trained in an auditory interval discrimination task during the sessions 2-6. In this task, a short burst of noise was presented on each trial (with the duration of the sound drawn from a uniform distribution in the range of 350-650 ms or 1100-1400 ms) and participants were asked to indicate whether it was shorter or longer than a reference duration that was presented numerically on the screen (e.g., "500 ms" or "1250 ms"). There were 700 trials in each session (for further details see Kattner, 2017).

Results

Due to extremely poor recall performance (two participants with $M < 3.5$ correctly recalled digits) or the absence of an irrelevant speech effect in the pre-test (four participants), the data of five participants in the Training group ($n = 18$; 15 women; $M_{age} = 22.6$ years) and one participant in the Control group ($n = 21$; 16 women; $M_{age} = 21.5$ years) were not included in the analysis.

For the Training group, both the median and the maximum number of letters reached for the target lists increased significantly from the first to the fifth training session [$F(4,68) = 7.84$; $p < .001$; $\eta^2_G = 0.10$ and $F(4,68) = 6.41$; $p < .001$; $\eta^2_G = 0.09$, respectively] (see Fig. 1A), indicating that the training with the dichotic listening task strengthened perceptual control and enhanced the selective encoding capacity for verbal information. Evidence of perceptual learning was also observed in the Control group who was trained on an auditory interval discrimination task with an initial discrimination threshold of 94.7 ms ($SD = 44.2$; range: 36.8 – 191.2 ms) in training session 1 and a final discrimination threshold of 75.9 ms ($SD = 32.9$; range: 27.9 – 143.96) in training session 5.

The irrelevant speech effect was determined for both the Training and the Control group before and after the five training sessions by subtracting the number of correctly recalled digits during speech (male or female) from the number of correctly recalled digits during noise. The resulting scores are illustrated in Fig. 1B for both visually presented and spoken target digits. At pre-test, the magnitude of the irrelevant speech effect did not differ between groups [$F(1,37) = 0.05$; $p = .83$], and there was no difference in recall performance between digits presented in the visual and auditory modality [$F(1,37) = 0.74$; $p = .39$]. However, a 2 (group: Training/Control) × 2 (modality: visual/spoken) × 2 (test: pre/post) mixed-factors ANOVA with modality and test as repeated-measures factors revealed a significant group × test interaction [$F(1,37) = 4.48$; $p = .04$; $\eta^2_G = 0.02$], indicating that the irrelevant speech effect was reduced in the Training group (from $M_{pre} = 1.08$; $SD_{pre} = 0.63$ to $M_{post} = 0.74$; $SD_{post} = 0.76$), but not in the Control group (from $M_{pre} = 1.04$; $SD_{pre} = 0.89$ to $M_{post} = 1.07$; $SD_{post} = 0.70$). There was no significant main effect of test [$F(1,37) = 2.69$; $p = .11$; $\eta^2_G < 0.01$], or modality [$F(1,37) = 1.30$; $p = .26$; $\eta^2_G = 0.01$], no three-way interaction between group, test, and modality, [$F(1,37) = 1.08$; $p = .31$; $\eta^2_G < 0.01$], and no other significant effects [$F < 1$]. Post-hoc analyses, however, revealed that the crucial group × test interaction was significant only for the trials with visually presented digits [$F(1,37) = 5.08$; $p = .03$; $\eta^2_G = 0.03$], but not for the trials with spoken digits [$F(1,37) = 0.41$; $p = .53$; $\eta^2_G < 0.01$].

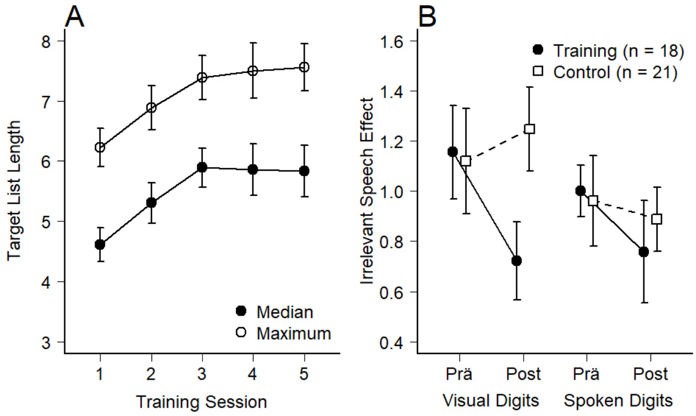


Figure 1. **(A)** Sample means for the median and maximum lengths of target lists to be encoded during the dichotic listening training across the five training sessions. **(B)** Irrelevant speech effects (difference in recall performance between trials with noise and speech) with visually presented and spoken target digits before (pre) and after (post) training for the Training and Control group. Error bars depict standard errors of the means.

Discussion and Conclusions

The present study demonstrates that a dichotic listening training effectively reduces the disruptions in serial recall performance produced by task-irrelevant speech by about 25%, whereas a control training with an auditory interval discrimination task did not affect the magnitude of the irrelevant speech effect. This finding suggests that the dichotic listening training influenced mechanisms of auditory attention or perceptual control that can be used to selectively attend to and encode relevant speech tokens while filtering out task-irrelevant speech (and thus avoiding perceptual interference). In addition to previous studies showing that the attentional capture effect of auditory distraction can be eliminated by cognitive control (Hughes et al., 2013), the present results show that the perceptual interference between task-irrelevant speech and the verbal seriation process can be reduced through a training of auditory perceptual control. Hence, in line with the duplex-mechanism account of auditory distraction (Hughes et al., 2007), it could be argued that attentional capture effect due to acoustical deviations may be susceptible to the cognitive control, whereas the perceptual interference of continuously fluctuating sound may be sensitive to perceptual control. Further research is required to test this assumption by (a) assessing whether the present perceptual control training also affects the attentional capture effect resulting from acoustical deviations, and (b) developing a cognitive control training that turns out to reduce or eliminate the attentional capture effect, but not the perceptual interference produced by continuous irrelevant speech.

Acknowledgements

We thank Corinne Orlemann and Celine Saul for their help with recruiting participants and running the experiments ($n = 8$ and $n = 13$ participants, respectively), and Katharina Rost for providing the speech recordings presented during the serial recall tasks.

References

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 443–446. http://doi.org/10.1163/156856897X00357

Hughes, R. W., Hurlstone, M. J., Marsh, J. E., Vachon, F., & Jones, D. M. (2013). Cognitive control of auditory distraction: impact of task difficulty, foreknowledge, and working memory capacity supports duplex-mechanism account. *Journal of Experimental Psychology. Human Perception and Performance*, *39*(2), 539–553. http://doi.org/10.1037/a0029064

Hughes, R. W., Vachon, F., & Jones, D. M. (2005). Auditory attentional capture during serial recall: Violations at encoding of an algorithm-based neural model? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(4), 736–749. http://doi.org/10.1037/0278-7393.31.4.736

Hughes, R. W., Vachon, F., & Jones, D. M. (2007). Disruption of short-term memory by changing and deviant sounds: Support for a duplex-mechanism account of auditory distraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(6), 1050–1061. http://doi.org/10.1037/0278-7393.33.6.1050

Kattner, F. (2017). Specificity of perceptual learning in an auditory interval categorization task. In K. Ueda, H. Ito, G. B. Remijn, & Y. Nakajima (Eds.), *Fechner Day 2017: Proceedings of the 33rd Annual Meeting of the International Society of Psychophysics* (p. 119). Fukuoka, Japan: International Society for Psychophysics.

Kattner, F., & Ellermeier, W. (2014). Irrelevant speech does not interfere with serial recall in early blind listeners. *Quarterly Journal of Experimental Psychology*, *67*(11). http://doi.org/10.1080/17470218.2014.910537

Kleiner, M., Brainard, D. H., Pelli, D. G., Broussard, C., Wolf, T., & Niehorster, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*(14), 1–16. http://doi.org/10.1068/v070821

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. http://doi.org/10.1163/156856897X00366

Salamé, P., & Baddeley, A. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior*, *21*, 150–164. http://doi.org/10.1016/S0022-5371(82)90521-7

PAUSE DURATION INFLUENCES IMPRESSIONS OF SPEECH STYLE IN ENGLISH PUBLIC SPEAKING

Shimeng Liu[1], Yoshitaka Nakajima[1], and Mark A. Elliott[2]
[1] *Department of Human Science, Faculty of Design,*
Kyushu University, Fukuoka, Japan
[2] *School of Psychology, National University of Ireland Galway, Galway, Ireland*
liushimengdl@163.com, nakajima@design.kyushu-u.ac.jp, mark.elliott@nuigalway.ie

In the present study we aim to find the most suitable pause duration in public speaking and to establish a more efficient, objective index for the training of public speaking. A listening experiment was carried out varying pause duration systematically, in which, both comma and period pauses were changed with the same duration (of 0.1, 0.2, 0.4, 0.8, 1.6, and 3.2 s in each stimulus). Fifteen students (10 male and 5 female) from Kyushu University Undergraduate and Graduate Schools were invited to join the experiment as participants. The method and analysis were based on Uchida (2005, The Japanese Journal of Educational Psychology). Amongst 20 evaluation items, principal-component analysis allowed 5 components to be extracted, and a varimax rotation led to two factors that were useful for judging the quality of speech. These were “naturalness of speech” and “speech rate”.

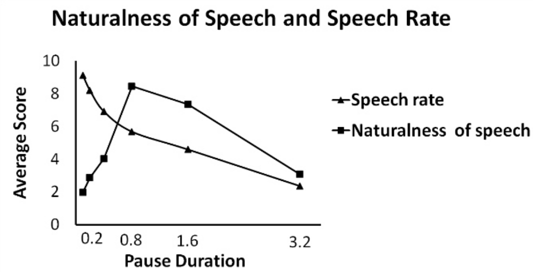


Fig. 1. The average factor scores regarding naturalness of speech and speech rate.

The first extracted factor comprised a component reflecting the "naturalness of speech". The items that comprised this scale were "naturalistic", "easy to understand", and "clear". The second factor concerned "speech rate". The items that comprised this scale were "fast-talking", "hasty", "long pause duration (negatively)". As a result and regarding of the naturalness of speech, speeches with pause durations from 0.8 to 1.6 s received the most positive evaluations. On the scale: “speech rate”, scores became lower when pause duration became longer. We anticipate our results will be of use in training speakers in ‘when’ and ‘how long’ to pause, and will allow speakers to become more conscious and confident in controlling the timing and rhythm of their speech. Our study offers strong scientific support for the fact that ideal speakers share the same properties: When giving a speech, ideal speakers leave a pause of suitable pause duration to help the audience feel the naturalness of their speech; meanwhile this also makes the speech rate suitable. Our results will be discussed alongside a second study in which we collected data from English native-speakers, with whom we sought to establish whether there are differences between the two groups.

EFFECTS OF UNCERTAINTY ON FEEDBACK PROCESSING IN A SPEECH-NONSPEECH DISCRIMINATION PARADIGM: BEHAVIORAL AND ELECTROPHYSIOLOGICAL EVIDENCE

Petra Ludowicy, Tina Weis, André Haese, Daniela Czernochowski, and Thomas Lachmann
Center for Cognitive Sxcience, University of Kaiserslautern, Germany
petra.ludowicy@sowi.uni-kl.de, lachmann @sowi.uni-kl.de

Effects of uncertainty during feedback processing were investigated in an auditory speech-nonspeech discrimination task. Participants assigned stimuli as either speech or non-speech. Stimuli were stepwise morphs of a speech sound (German vowels /a/ or /a:/) and its spectrally rotated counterpart, which were chosen as levels of uncertainty depending on the morph ratio. Positive, negative and uninformative performance feedback was given. In a second experiment, participants rated these stimuli on a 6-point scale ranging from "certainly speech" to "certainly non-speech". Behavioral data and event related potential (ERP) components anterior N1, feedback related negativity (FRN) and P3 were analyzed. Amplitudes of all components were larger for negative compared to positive feedback. Regarding uninformative feedback, anterior N1 and FRN were similar for negative and uninformative feedback. P3 amplitudes for uninformative feedback were just in-between those of positive and negative feedback. Contrarily to prior studies, uncertainty affected only P3 in terms of larger amplitudes compared to the certain condition.

SOUND ASSESSMENT OF AIRCRAFT OVERFLIGHTS

Svea Missfeldt and Rainer Höger
Institute of Experimental and Industrial Psychology,
Leuphana University, D-21335 Lüneburg, Germany

In an environment of constant growth and omnipresent soundscapes it is almost redundant to say that noise pollution will take on an even stronger role in future challenges. Noise research has found negative effects not only on sleep quality and physical health it has also shown that aircraft noise has a negative effect on the reading ability of school children located in the airport neighborhood. Therefore, alongside eye movement recordings while reading under different noise conditions (analysis in progress), we retrieved sound assessment data regarding the pleasantness and arousal perceived when being exposed to different noises. Results show that an aircraft overflight approaching from the reader's right ear, moving contrary to the reading direction, is perceived as significantly more unpleasant than an aircraft approaching from the left side moving congruently. This might indicate that the moving information implied in aircraft overflight noise is causal to the negative impact. We believe that the attention capturing effect of the moving noise source occupies mental capacity and affects the inner attention which is engaged in the reading process.

AN ACOUSTIC ANALYSIS OF PREPOSITION PHRASES IN ENGLISH

Xiaoyang Yu[1], Yoshitaka Nakajima[2], Yixin Zhang[1], Takuya Kishida[3], and Kazuo Ueda[2]
[1] *Graduate School of Design, Kyushu University, 4-9-1 Shiobaru Minani-ku, Fukuoka, 815-8540 Japan*
[2] *Dept. Human Sciences/Research Center for Applied Perceptual Science, Kyushu University, 4-9-1 Shiobaru Minami-ku, Fukuoka, 815-8540 Japan*
[3] *Department of Human Science, Kyushu University, 4-9-1 Shiobaru Minami-ku Fukuoka, 815-8540 Japan*
yuxiaoyangat123@gmail.com, nakajima@design.kyushu-u.ac.jp, zhangyixin0210@yahoo.com, kishida.takuya0119@gmail.com, ueda@design.kyushu-u.ac.jp

The present study was on how spectral-change factors behave in English preposition phrases consisting of prepositions and noun phrases. Spectral changes of spoken English sentences in our new database were subjected to origin-shifted factor analysis. Three factors as in our previous study [Kishida et al. (2016). Front. Psychol. 7:517] appeared. One of the factors, the *mid-low-frequency factor* was closely related to a frequency range around 1100 Hz, and another factor, the *high-frequency factor* was closely related to a frequency range above 3300 Hz. The *high-frequency factor* scores were higher in the noun phrases than in prepositions, but the factor score difference of the *mid-low-frequency factor* was not clear. Frequency components above 3300 Hz may play important roles to clarify noun phrases perceptually.

MULTIVARIATE ACOUSTIC ANALYSIS OF INITIAL CONSONANT CLUSTERS IN ENGLISH

Yixin Zhang[1], Yoshitaka Nakajima[2], Xiaoyang Yu[1], Kazuo Ueda[2], Takuya Kishida[3], Gerard B. Remijn[2], Sophia Arndt[4], and Mark A. Elliott[4]
[1] *Graduate School of Design, Kyushu University,*
4-9-1 Shiobaru Minani-ku, Fukuoka, Japan,
[2] *Dept. Human Sciences/Research Center for Applied Perceptual Science,*
Kyushu University, 4-9-1 Shiobaru Minami-ku, Fukuoka, Japan
[3]*Department of Human Science, Kyushu University,*
4-9-1 Shiobaru Minami-ku Fukuoka, Japan
[4]*School of Psychology, National University of Ireland Galway,*
Galway, Republic of Ireland
zhangyixin0210@gmail.com, nakajima@design.kyushu-u.ac.jp

Two or more consonants adjacent to one another within a syllable make a consonant cluster. To pronounce and perceive consonant clusters appropriately is important for learners of English as a foreign language. In order to support such effort, consonant clusters at the initial positions of English syllables were analyzed together with the following vowels in terms of spectral changes. A new database of spoken English sentences was created for this purpose. The newly recorded sentences were represented as power fluctuations in 20 critical-band filters simulating auditory periphery, and the power fluctuations (the temporal changes of 20 variates) were subjected to factor analysis. The analysis method was modified so that the extracted subspace always would contain the point representing acoustic silence. In the 3-factor analysis, a factor closely related to frequency components in a range around 1100 Hz and another factor related to components near or above 3300 Hz appeared. These two factors were exclusive: If the factor score of one of these factors was high, then the factor score of the other was close to zero, forming an L-shaped scattergram as in our previous study [Nakajima et al. (2017). Sci. Rep., 7: 46049]. If an initial consonant cluster and the following vowel were to be represented in this factor space, the spectral change also had to be bound within this L- shape. This was confirmed for our database, and the 3-dimensional motion representing the spectral change was analyzed further

Reference

Nakajima, Y., Ueda, K., Fujimaru, S., Motomura, H. & Ohsaka, Y. (2017). English phonology and an acoustic language universal. *Scientific Reports, 7: 46049.*

POSTER SESSION II

GENERALIZATION ABILITIES OF PREVALENT DEEP GENERATIVE MODELS AND THE HUMAN BRAIN: A PSYCHOPHYSICS PERSPECTIVE

Rajesh Chidambaram[1]*, Vera Eymann[1], Michael Kampffmeyer[2]*, Xiaodan Liang[3], and Thomas Lachmann[1,4]

[1] *Center for Cognitive Science, TU Kaiserslautern*

[2] *UiT the Arctic University of Norway, Tromsø*

[3] *Machine Learning Department, Carnegie Mellon University*

[4] *Department of Experimental Psychology, University of Leuven*

* *Visiting Scholar, CMU, Pittsburgh*

rajesh1990@live.in

Abstract

Machine learning algorithms have recently matched human-level performances in various domains including object detection, machine translation, and video game playing. Deep learning, a sub-class of Machine Learning, attempts to mimic human-level cognitive abilities. More specifically, for the task of text-to-image synthesis, zero-shot learning abilities have been demonstrated using variants of Generative Adversarial Networks (GANs). Yet, their ability to generalize concepts for such generations is experimental. In this work, we analyze the similarities and differences between human-level cognition and prevalent deep generative models by utilizng perspectives on geometric generalization tasks from psychophysics. By increasing the complexity of the visual stimulus, we analyze the performace of text-to-image synthesis algorithms. We conclude by establishing that, an interdisciplinary approach to machine learning is required to achieve human-level generalization abilities.

Advancements in Machine Learning have enabled silicon-based intelligent systems match human-level performances in a variety of tasks such as object detection, machine translation, and video game playing (LeCun, Bengio, & Hinton, 2015). Yet, generalizing concepts across classes/domains remain a key challenge for such algorithms. Geometric generalization offers a standardized platform to test and evaluate generalization abilities of Machine Learning algorithms. Initial attempts have been made to implement neurodynamic models that can generalize geometrical concepts from a fewer visual representations (Alexander, Trengove, Sheridan, & van Leeuwen, 2011). In a previous study (Jaarsveld, Lachmann, Hamel, & Leeuwen, 2010), it was shown that humans can not only solve geometry based intelligent tests, but also create one. But state-of-the-art Machine Learning algorithms can neither solve nor generate such intelligence tests without being explicitly programmed.

For biological brains, Gestalt laws provide the '*principles of grouping*' and describes how the visual system naturally perceives objects (Wertheimer, 1923). Structural Information Theory (SIT), a Gestalt approach to plausible perceptual interpretations provides a mathematical formulation via quantitative models (Leeuwenberg & Van der Helm, 2013). Though Gestalt laws elaborates on the task of object recognition, they do not elaborate on how human beings construct mental models from mere textual inputs. Whereas, the human brain comprehends textual contents by constructing mental models (Woolley, 2011). Since the permutations and combinations of all semantically plausible textual inputs in a language is infinite by nature, it is not possible to construct mental models for all such inputs without generalization of concepts. Hence, to better understand human cognition and improve the state-of-the-art Machine Learning algorithms, a detailed study of models that generate images from mere textual inputs is essential.

In the present study, we analysed the differences between human-level cognition and prevalent deep generative methods from psychophysics perspectives. By increasing the complexity of the visual stimulus, we analysed the accuracy of text-to-image synthesis algorithms and henceforth their ability to generalize.

Geometric Generalization

Generalization can be defined as the ability of an agent to learn upon training and then apply its learning to cases the agent had never seen before (Cohn, Atlas, & Ladner, 1994). Because generalization plays a key role in fluid intelligence, it is part of intelligence tests such as the Cattell Culture Fair IQ test (Cattell, 1963) and the Progressive Matrices (Raven & Court, 1998).

Geometric generalization is defined as the ability of a learning agent to generalize the concept of polygons to infinite numbers, after being introduced to lower-order polygons only (Chidambaram et al., 2018). Geometric generalization cannot be performed via rote-memorization. To generalize concepts across classes, it involves understanding the concept of polygons through abstract reasoning. Unlike other forms of generalization, geometric generalization is more precisely evaluable and culturally independent.

Method

Variants of Generative Adversarial Networks (GANs) are widely adopted generative models in Machine Learning. To perform unsupervised learning on a supervised loss, a generator neural network G and a discriminator neural network D are simultaneously trained to outperform each other (Goodfellow et al., 2014). This is achieved by cleverly combine the min-max problem to the loss function of the composite deep neural network as shown below:

$$\min_{G}\max_{D} V(D,G) = \mathbb{E}_{x\sim p_{data}}[logD(x)] + \mathbb{E}_{z\sim p_z}[log(1-D(G(z)))]$$

In the above equation, *p-data* indicates the probability distribution of real data. *p-z* indicates the probability distribution from which a latent variable is chosen and fed to the generator G. To perform text-to-image synthesis, GANs are conditioned with textual inputs at both Generator G and discriminator D stages. Upon training, text-to-image synthesis GANs are able to exhibit zero-shot learning by generating realistic images on unseen textual inputs. However, their ability to generate unseen texts through human-level abstract reasoning are experimental. Research on Generative Adversarial Text-to-Image synthesis (GAT2I) (Reed et al., 2016) and Attentional Generative Adversarial Networks (Attn-GAN) (Xu et al., n.d.) show that they are unable to perform at human level accuracies (Chidambaram et al., 2018) on the dataset '3-9 *world*'. As mentioned earlier, in this study, we perform further analysis from psychophysics perspectives by increasing the complexity of the geometric figures. GAT2I and Attn-GAN were trained on the dataset '3-9 *world*', a subset of the dataset '*Infinite World*'. '*Infinite World*' poses geometric generalization based tests for Machine Learning algorithms. Analogous to the Zero-Shot Intelligence (ZSI) scoring method for the tests proposed in the '*Infinite world*' dataset, accuracy was calculated by squashing partially correct answers between 25 to 75.

We trained our models on *NVIDIA-Titan* GPUs. We used an open-sourced version of GAT2I from *GitHub*. This version of GAT2I uses skip-thought vectors for word embeddings and the TensorFlow framework for its implementation. The open-sourced Attn-GAN model we used was implemented on the PyTorch framework.

Geometric generalization abilities of 'Generative Adversarial Text-to-Image Synthesis'

As shown in the figure (Fig. 1), we computed the average accuracy (upto 100 epochs) of the generated geometric figures while increasing the complexity of the stimuli. Here, complexity refers to the number of sides in the polygon. For higher order polygons, decreasing levels of performances in humans might be attributed to the lack of motivation for drawing complex figures. When the number of sides in the polygon were increased, a slight decrease in the accuracy levels were observed. Interestingly, GAT2I performs better on regular polygons than on irregular polygons. It is notable that regular polygons can be generated via rote memorization owing to their regular structure. But irregular polygons cannot be memorized since they can take highly random shapes.

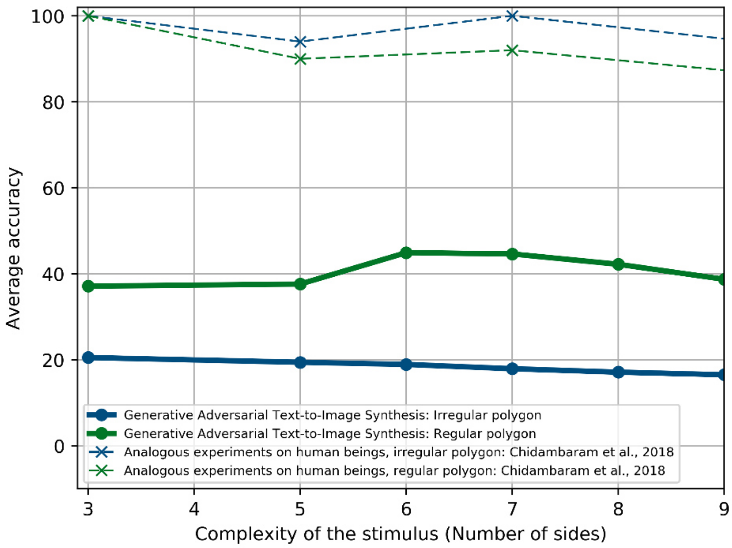


Fig. 1. Accuracy of 'Generative Adversarial text-to-image synthesis' on increasing complexity (number of sides) of regular and irregular polygons.

Geometric generalization abilities of 'Attentional Generative Adversarial Networks'

Attn-GAN uses three GANs for generating high resolution images from textual inputs. Upon performing feature extraction, the output of the first GAN is fed as input to the second GAN and so-forth. In addition, the word/sentence embeddings are conditioned at different stages of the GANs. We chose this model to analyze the effects of recurrent attention on GANs. For any given complexity of the polygon, as shown in the figure (Fig. 2), the performance of Attn-GAN is not comparable to that of the humans.

Interestingly, the performance of Attn-GAN improves upon increasing complexity. This is in sharp contrast with that of human-beings, whose performances tend to decrease on increasing complexity of the visual stimulus. This might explain that the mechanisms in which the visual information is processed differs from the two systems. This could also possibly explain the failure of Attn-GAN on geometric generalization tasks. Since deep neural network methods continue to remain uninterpretable, it is difficult to mathematically explain why Attn-GAN performs better when the number of sides in the polygon are increased.

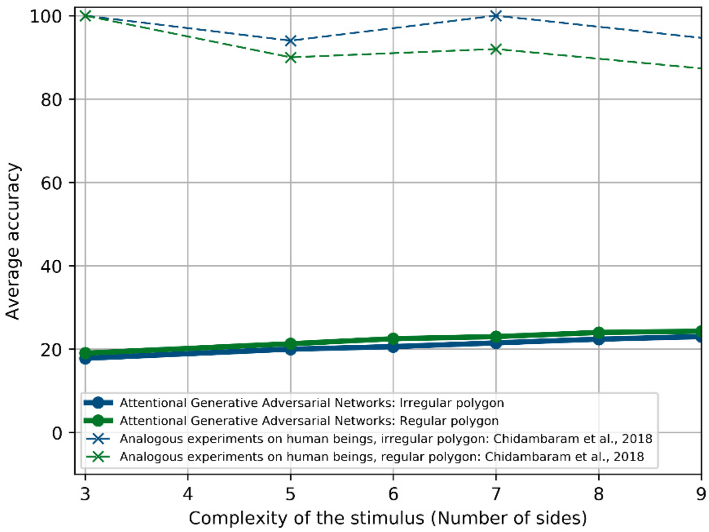


Fig. 2. Accuracy of 'Attentional Generative Adversarial Networks' on increasing complexity (number of sides) of regular and irregular polygons.

Generalizing Generalization

Artificial Intelligence and neuroscience shares a long, intertwined history (Hassabis, Kumaran, Summerfield, & Botvinick, 2017). Yet, the two fields have recently lost communication (Marblestone, Wayne, & Kording, 2016). A unified, interdisciplinary approach is required to better understand the generalization abilities of the human brain and thereby develop General Artificial Intelligence. Hence, in an attempt to improve the generalization abilities of Machine Learning algorithms, we attempt to unify the concept of generalization. The ability to generalize concepts across classes/domains is closely related to a range of disciplines such as few-shot learning (Snell, Swersky, & Zemel, 2017), meta-cognition (Borkowski, Estrada, Milstead, & Hale, 1989; Zohar & Ben David, 2009), domain adaptation (Motiian, Piccirilli, Adjeroh, & Doretto, 2017; Zhang, Zhang, & Ye, 2012), transfer learning (Bengio, 2012; Lampert, Nickisch, & Harmeling, 2009; Yao & Doretto, 2010), non-convex optimization (Kesavan & Barton, 2000; Takeda & Sugiyama, 2009), Bayesian inference(Tenenbaum & Griffiths, 2001; Tenenbaum, Kemp, Griffiths, & Goodman, 2011; Xu & Tenenbaum, 2007), perception (Fahle, 2005; Liu & Weinshall, 1999; Sireteanu & Rettenbach, 2000), creativity, (Ervynck, 2002; Horzyk, 2014; Plucker, 2004) and inductive, and abstract reasoning (Heit, 2000; Krueger & Clement, 1996). This is diagramatically represented in Fig. 3. To perform zero-shot or few-shot learning, it might be necessary to actively unlearn the perceptually learrned components and combine concepts from different domains in a creative manner through abstract reasoning. To deliberately learn a representation, it requires meta-level cognition over the prior information. Depending upon the complexity of the generalization task in hand, the model should either adopt passive perceptual learning or active transfer-learning.

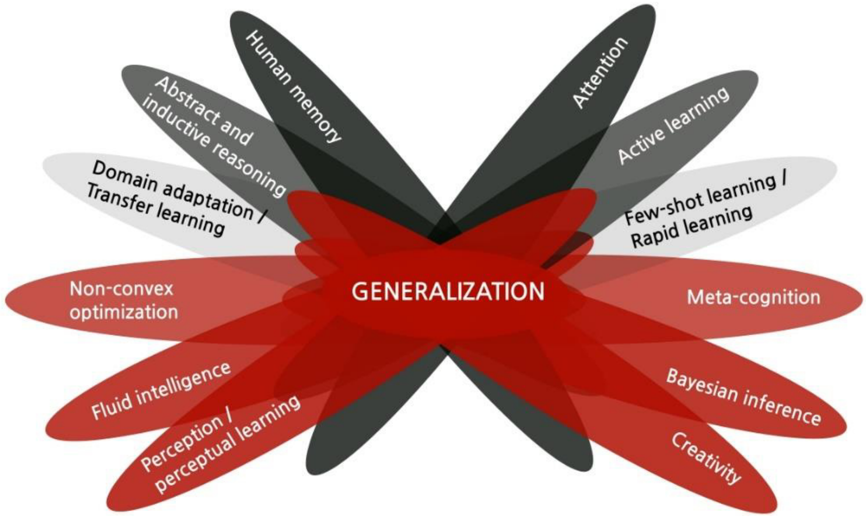


Fig. 3. The concept of generalization can itself be generalized from a range of scientific disciplines such as Machine Learning, human-cognition, mathematics for non-convex optimization, cognitive neuroscience, and philosophy of mind.

Conclusion

In this work, we analysed the performance levels of GANs on geometric generalization tasks while varying the complexity of the stimuli. Our analysis on GAT2I indicates that, increasing the complexity of the stimulus by increasing the number of sides in the polygon from 3 to 9, slightly decreases the performance levels. On the other hand, Attn-GAN performs slightly better as the complexity of the visual stimulus increases. This is in sharp contrast to the way in which humans perform. Confirming previous studies, unlike humans, the GAN models are clearly unable to generalize the concept of polygons. Both GAT2I and Attn-GAN performs comparatively better on texts for regular polygons than for irregular polygons. This further evinces that the current state-of-the-art GANs rely on rote-memorization and not on generalizing geometric concepts.

Generalization as we showed, is an integration of concepts from several disciplines of science. Therefore, we can conclude that, for machine learning algorithms to solve the problem of generalization, we will need a deeper understanding of what type of generalization the task demands and henceforth adopt techniques that will suite the purpose.

Acknowledgements

The authors thank Radha Nila Meghanathan for proof reading an earlier version of the paper. The authors thank Malizia Kupper of the Fechner day'18 organising committee for supporting us in submitting the current version of the manuscript.

References

Alexander, D. M., Trengove, C., Sheridan, P. E., & van Leeuwen, C. (2011). Generalization of learning by synchronous waves: from perceptual organization to invariant organization. *Cognitive Neurodynamics*, *5*(2), 113–132.

Bengio, Y. (2012). Deep learning of representations for unsupervised and transfer learning. In *Proceedings of ICML Workshop on Unsupervised and Transfer Learning* (pp. 17–36).

Borkowski, J. G., Estrada, M. T., Milstead, M., & Hale, C. A. (1989). General problem-solving skills: Relations between metacognition and strategic processing. *Learning Disability Quarterly*, *12*(1), 57–70.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1.

Chidambaram, R., Kampffmeyer, M., Neiswanger, W., Liang, X., Lachmann, T., & Xing, E. (2018). *Geometric Generalization Based Zero-Shot Intelligence Metric*.

Cohn, D., Atlas, L., & Ladner, R. (1994). Improving generalization with active learning. *Machine Learning*, *15*(2), 201–221. https://doi.org/10.1007/BF00993277

Duncan, J., Burgess, P., & Emslie, H. (1995). Fluid intelligence after frontal lobe lesions. *Neuropsychologia*, *33*(3), 261–268.

Ervynck, G. (2002). Mathematical creativity. In *Advanced mathematical thinking* (pp. 42–53). Springer.

Fahle, M. (2005). Perceptual learning: specificity versus generalization. *Current Opinion in Neurobiology*, *15*(2), 154–160.

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning, vol. 1. MIT press Cambridge.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., … Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672–2680).

Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron*, *95*(2), 245–258.

Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review*, *7*(4), 569–592.

Horzyk, A. (2014). How does generalization and creativity come into being in neural associative systems and how does it form human-like knowledge? *Neurocomputing*, *144*, 238–257.

Jaarsveld, S., Lachmann, T., Hamel, R., & Leeuwen, C. van. (2010). Solving and creating Raven Progressive Matrices: Reasoning in well-and ill-defined problem spaces. *Creativity Research Journal*, *22*(3), 304–319.

Kesavan, P., & Barton, P. I. (2000). Generalized branch-and-cut framework for mixed-integer nonlinear optimization problems. *Computers & Chemical Engineering*, *24*(2–7), 1361–

1366.
Krueger, J., & Clement, R. W. (1996). Inferring category characteristics from sample characteristics: Inductive reasoning and social projection. *Journal of Experimental Psychology: General*, *125*(1), 52.
Lampert, C. H., Nickisch, H., & Harmeling, S. (2009). Learning to detect unseen object classes by between-class attribute transfer. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on* (pp. 951–958). IEEE.
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.
Leeuwenberg, E., & Van der Helm, P. A. (2013). *Structural information theory: The simplicity of visual form*. Cambridge University Press.
Liu, Z., & Weinshall, D. (1999). Mechanisms of generalization in perceptual learning. In *Advances in Neural Information Processing Systems* (pp. 45–51).
Marblestone, A. H., Wayne, G., & Kording, K. P. (2016). Toward an integration of deep learning and neuroscience. *Frontiers in Computational Neuroscience*, *10*, 94.
Motiian, S., Piccirilli, M., Adjeroh, D. A., & Doretto, G. (2017). Unified deep supervised domain adaptation and generalization. In *The IEEE International Conference on Computer Vision (ICCV)* (Vol. 2).
Plucker, J. A. (2004). Generalization of creativity across domains: Examination of the method effect hypothesis. *The Journal of Creative Behavior*, *38*(1), 1–12.
Sireteanu, R., & Rettenbach, R. (2000). Perceptual learning in visual search generalizes over tasks, locations, and eyes. *Vision Research*, *40*(21), 2925–2949.
Raven, J. C., & Court, J. H. (1998). *Raven's progressive matrices and vocabulary scales*. Oxford Psychologists Press.
Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., & Lee, H. (2016). Generative adversarial text to image synthesis. *ArXiv Preprint ArXiv:1605.05396*.
Snell, J., Swersky, K., & Zemel, R. (2017). Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems* (pp. 4080–4090).
Takeda, A., & Sugiyama, M. (2009). On Generalization Performance and Non-Convex Optimization of Extended v-Support Vector Machine. *New Generation Computing*, *27*(3), 259–279. https://doi.org/10.1007/s00354-008-0064-6
Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, *24*(4), 629–640.
Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285.
Wertheimer, M. (1923). A brief introduction to gestalt, identifying key theories and principles. *Psychol Forsch*, *4*, 301–350.
Woolley, G. (2011). Reading comprehension. In *Reading Comprehension* (pp. 15–34). Springer.
Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245.
Xu, T., Zhang, P., Huang, Q., Zhang, H., Gan, Z., Huang, X., & He, X. (n.d.). Attngan: Fine-grained text to image generation with attentional generative adversarial networks.
Yao, Y., & Doretto, G. (2010). Boosting for transfer learning with multiple sources. In *Computer vision and pattern recognition (CVPR), 2010 IEEE conference on* (pp. 1855–1862). IEEE.
Zhang, C., Zhang, L., & Ye, J. (2012). Generalization bounds for domain adaptation. In *Advances in neural information processing systems* (pp. 3320–3328).
Zohar, A., & Ben David, A. (2009). Paving a clear path in a thick forest: a conceptual analysis of a metacognitive component. *Metacognition and Learning*, *4*(3), 177–195. https://doi.org/10.1007/s11409-009-9044-6

VALENCE OF TARGETS AND OF FLANKERS AFFECTS CRITICAL SPACING IN CROWDING

Lisa V. Eberhardt, Anke Huckauf, and Ferdinand Pittino
Ulm University, Ulm, Germany
prename.surname@uni-ulm.de

A stimulus presented in the visual periphery is prone to interference from adjacent stimuli. This so-called crowding effect is the more pronounced the further the target is displayed in the periphery and the smaller the spacing between target and flankers, suggesting early visual factors underlying crowding (e.g., Bouma, 1970; Levi, 2008). However, there are already indications that higher-level information of target and flankers is processed even under crowded conditions (e.g., Huckauf & Heller, 1999; Faivre, Berthet, & Kouider, 2012). Affective information is assumed to be prioritized. In order to examine effects of prioritized processing on crowding, we made use of evaluative conditioning. Here, stimuli (in our case Landolt rings) are repeatedly paired with affective stimuli (in our case negative and neutral pictures of the International Affective Picture System, IAPS; Lang, Bradley, & Cuthbert, 2008). After conditioning, conditioned Landolt rings served as targets and as flankers in a visual crowding task. We measured the 75%-threshold over spacing by the adaptive Bayesian QUEST function (Watson & Pelli, 1983). The results show larger critical spacing for negative than for neutral flankers. For conditioned targets, the expected smaller critical spacing for negative targets was weaker and resulted in a marginal effect (Pittino et al., submitted). We replicated this effect in three-dimensional space, when crowded stimuli (either targets or flankers) were presented in front of or behind the fixation plane. This was realized by presenting target and flankers separately on two screens, which are placed at different distances to the observer and are simultaneously viewed through a semi-transparent mirror (see Eberhardt & Huckauf, 2017). The study showed that evaluative information affects crowding even with defocused stimuli. The data show that evaluative conditioning leads to prioritized processing of flankers as well as of crowded targets.

WHERE DO WE LOOK NEXT? FIRST STEPS TOWARDS A SCAN PATH THEORY

Rainer Höger and Marlene Schlorf
Institute of Experimental Industrial Psychology, Leuphana Universität Lüneburg, Wilschenbrucher Weg 84, 21335 Lüneburg/Germany
hoeger@leuphana.de

Abstract

While viewing pictures a series of eye-movements (saccades) were performed which constitute the so-called scan path. The ideas of how those sequences of fixations are steered by the cognitive system range from saliency based mechanisms up to global and local scanning strategies. The study reported follows the idea that spatial frequency components of image parts influence the sequence of fixations. Within an experiment sinusoidal gratings of different spatial frequencies (0.4 cpd to 7 cpd) were arranged within an image. A series of images in which the position of the different gratings were arranged randomly were shown to 44 participants. Eye-movements were recorded while recipients viewed these compound images. The gaze data were analyzed with regard to the fixation sequence of the different gratings. The results showed that participants tend to fixate gratings with lower spatial frequencies (around 1 cpd) at first. The second glance is merely directed to gratings with much lower spatial frequencies (0.5 cpd) whereas in the following third glance the recipients returned near to the initially inspected frequency band. This finding can be interpreted as a perceptual strategy to inspect at first more global structures. Moreover, before the initial frequency band is visited again another frequency band is sought out. This observation looks like a spatial frequency related version of the inhibition of return effect.

During viewing an image the observer shifts his gaze across the scene. The underlying rationale is that humans direct the region of their retina with the highest resolution (fovea) to aspects of the optical scene which are of a certain relevance for the observer (Nelson and Loftus, 1980). The steering process of those gaze shifts can be divided into two aspects: stimulus-based and knowledge-based mechanisms (Henderson, 2003). Stimulus-based processes are those which refer to graphical aspects of an image such as color, contrast or edges (Richards and Kaufmann, 1969; Mannan et al., 1996). More elaborated stimulus-based models use saliency-maps to predict fixation positions. Saliency-maps refer to low level processing of the primary visual cortex where low level features such as orientation of lines, luminance changes (Itti and Koch, 2001; Torralba, 2003) and spatial frequency components (Reinagel and Zador 1999) were extracted from an image. Knowledge based gaze control refers to top-down processes where the meaning of image parts determines the positioning of fixations (Henderson et al., 1999).

An intermediate position between stimulus- and knowledge-based scan path mechanisms is taken by concepts which emphasize strategic aspects of fixation sequences. Under certain conditions fixations start at global stimulus attributes whereas under changing conditions the fixation sequence proceeds on a local level (Zangemeister et al., 1995).

The current study focuses on the role of spatial frequencies as stimulus-based aspects influencing the location of fixations. It is also investigated of how spatial frequency components are involved in strategic aspects referring to the serial order of fixations.

Method

Participants

44 participants took part in the experiment as unpaid volunteers. They were aged between 19 and 36 years; 15 were male and 29 female. All participants had normal visual acuity.

Material and Procedure

The stimuli used in the experiment consisted of 8 different sinusoidal gratings with spatial frequencies of 0.4, 0.5, 0.75, 1, 2, 4, 5, and 7 cycles per degree (cpd). Figure 1 shows exemplarily three gratings with spatial frequencies of 0.4, 1, and 2 cpd. All gratings were adjusted in contrast by using the standard contrast sensitivity function (Campbell and Maffei, 1974). In addition, the overall brightness of each grating was adjusted to the same value.

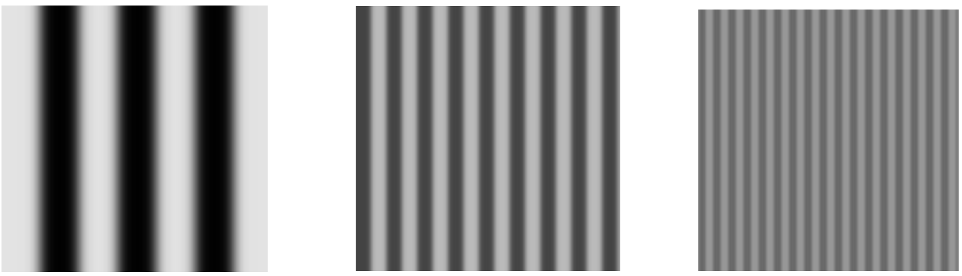


Fig. 1. Examples of adjusted sinusoidal gratings used in the experiment.

For the actual experiment the 8 different gratings were arranged to a compound stimulus as depicted in Figure 2.

The compound stimulus was presented on a video screen (48 cm x 27 cm) while the eye-movements of the participants were recorded via a high-speed eye tracking system (SMI Hi Speed 1250 Hz). The distance between the participant's eyes and the video screen was set to 60 cm obtaining the exact spatial frequency values. To prevent a confounding between spatial position and spatial frequency of a grating within the compound stimulus, 8 different variants of the compound were generated. The spatial distribution of the different gratings was counterbalanced using a Latin square procedure. The different variants then were presented to the participants in a random order. Each compound stimulus was presented for 5 seconds whereby each trial started with a fixation cross at the center of the screen. The participants were instructed first to fixate the cross in the middle and then look around without any special intention.

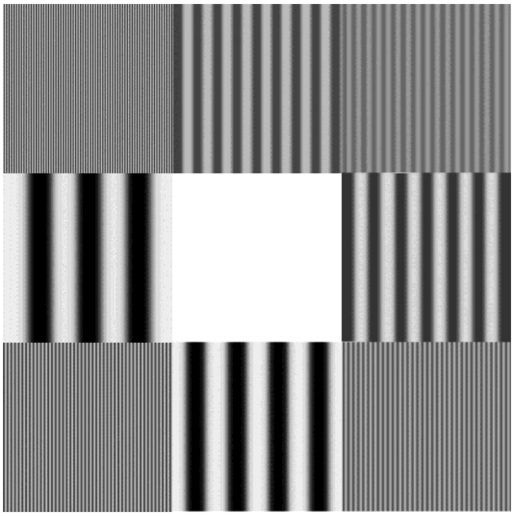


Fig. 2. Compound stimulus of gratings with spatial frequencies ranging from 0.4 up to 7 cpd.

Results and Discussion

For analyzing the sequence of fixations in a first step each grating within the compound stimulus was defined as a separate area of interest (AOI). The definition of AOIs is helpful for analyzing the fixation data semi-automatically using a corresponding analysis program (e.g. SMI BeGaze). Each fixation which falls into an AOI is counted independently of its exact position within the AOI. Figure 3 depicts an example of a fixation sequence of a recipient for one of the presented compound stimuli.

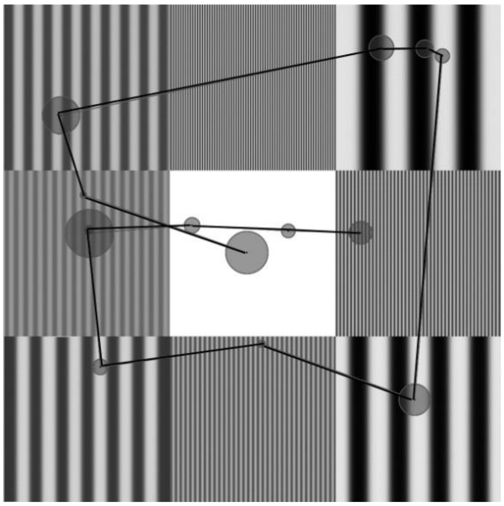


Fig. 3. Scan path of a recipient while viewing the compound for 5 seconds. The diameters of the grey circles reflect the fixation durations.

In the next step of analysis for each of the gratings within the compounds the frequency of the first fixation was counted. In the following step the same procedure was applied for the second fixation and afterwards for the third fixation. The results of this frequency statistic are shown in Figure 4.

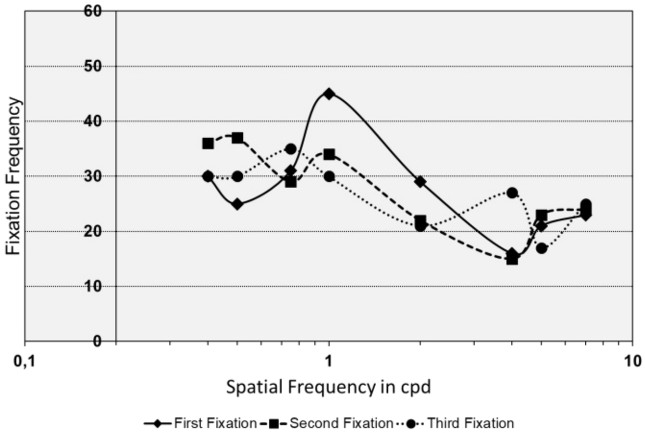


Fig. 4. Fixation distribution for the first, second, and third fixation

All three depicted distribution curves show fixation preferences for different spatial frequencies. As can be seen in Figure 4, for the first fixation the gaze is most frequently directed to the 1 cpd grating. The smallest number of fixations receives the grating with a spatial frequency of 4 cpd. The second fixation has its maximum around 0.5 cpd and its minimum at a spatial frequency of 4 cpd. The third fixation is most frequently directed to the 0.75 cpd grating and least frequently to the 5 cpd grating. Whether the three different curves for the first, second, and third fixation differ significantly from an equal distribution of fixations was checked by a CHI^2-test. For the first and the second fixation curve significant deviations from an equal distribution were obtained ($CHI^2 1stfix = 14{,}47$, $df = 7$, $p < .05$; $CHI^2 2ndfix = 15{,}51$, $df = 7$, $p < .05$) The curve for the third fixation failed to reach a significant deviation from equally distributed fixation frequencies ($CHI^2 3rdfix = 8{,}6$, $df = 7$, n.s.).

Taken together the findings can be interpreted as follows: the first glance is captured by a grating whose spatial frequency is lower than the spatial frequency area to which the human visual system is particularly sensitive (cp. Campbell and Maffei, 1974). Possibly the preference for lower spatial frequencies around 1 cpd reflects the tendency to examine global aspects within an image at first. The second fixation seems to enlarge this low frequency region towards deeper spatial frequencies (around 0.5 cpd). Within the first three fixations preferences for higher spatial frequencies above 4 cpd are considerably less pronounced. Insofar the idea that the visual system starts with collecting global information receives further support. Within the literature the preferred processing of low frequency spatial information has already been discussed in detail (Höger, 2001).

For the third fixation the preferences are kept at a low level across all spatial frequencies. This might mean that with the ongoing scanning process the preference for certain spatial frequencies is lowered. An inspection of the three different preference curves show maxima at different positions of the spatial frequency scale.

The answer to the question ‘where do we look next’ is that those locations are fixated whose spatial frequencies have not been processed before. This scanning strategy recalls the concept of ‘inhibition of return’ (Posner, 1985) in the domain of attention theories.

References

Campbell, F. W., & Maffei, L. (1974). Contrast and spatial frequency. *Scientific American*, 231, 106-115.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, 7, 498-504.

Henderson, J. M., Weeks, P. A., & Hollingworth, P. A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 210-228.

Höger, R. (2001). *Raumzeitliche Prozesse der visuellen Informationsverarbeitung*. Magdeburg: Scriptum.

Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Review Neuroscience*, 2, 194-203.

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Vision Research*, 10, 165-188.

Nelson, W. W., & Loftus, G. R. (1980). The functional visual field during picture viewing. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 391-399.

Posner, M. I., Rafal, R. D., Choate, L. S., & Vaughan, J. (1985). *Cognitive Neuropsychology*, 2, 211-228.

Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network; Computation in Neural Systems*, 10, 341-350.

Richards, W., & Kaufmann, L. (1969). ‘Center-of-gravity’ Tendencies for fixations and flow patterns. *Perception and Psychophysics*, 5, 81-84.

Torralba, A. (2003). Modeling global scene factors in attention. *Journal of the Optical Society of America A*, 20, 1407-1418.

Zangemeister, W. H., Sherman, K., & Stark, L. (1995). Evidence for a global scanpath strategy in viewing abstract compared with realistic images. *Neuropsychologia*, 33, 1009-1025.

ENHANCING THE COGNITIVE AND MOTOR ABILITIES OF VERY YOUNG CHILDREN: A PILOT STUDY OF THE EFFICACY OF THE PLAYWISELY APPROACH

Craig Leth-Steensen[1], Elena Gallitto[2], and Patricia A. Hannan[3]
[1] *Carleton University, Ottawa, Canada*
[2]*University of Ottawa, Ottawa, Canada*
[3] *PlayWisely, Dallas, TX, USA*
Craig.LethSteensen@carleton.ca, egall065@uottawa.ca, hannanpatty3@gmail.com

Abstract

PlayWisely is a novel approach to early learning designed to target the positive development of a wide range of cognitive and physical/motor abilities by stimulating the rapidly developing brain of very young children (from 6 months to 3 years of age). The current pilot study represents the first step towards providing an evidential basis for the efficacy of this approach by conducting a small-sample (N = 19) randomized controlled trial (RCT) comparison of the cognitive and motor abilities of children who were administered 16 weeks of PlayWisely training with children who were administered this training 5 months later. Results showed a marginally significant differential increase in cognitive scale scores (as measured by the Merrill-Palmer-Revised Scales of Development) over the 10-month study period.

Most organizations, educators, and researchers whose focus is on child development (e.g., National Scientific Council on the Developing Child; see also the Special Section "Laying the Foundation for Lifetime Learning" in *Science*, 2011, pp. 951-983) now recognize the vast potential that very early childhood education has with respect to enhancing developmental outcomes throughout the lifespan. However, the presence of actual programs to realize enhanced cognitive and motor development through psycho-educational training at such young ages (i.e., from infancy on) is severely lacking.

In this vein, a pilot study was conducted to gauge the potential efficacy of a novel approach to early learning called PlayWisely (Kern et al., 2013). This program is the first comprehensive early learning method to include a multitude of fun activities specifically designed to foster both cognitive and motor skills starting from a very young age. Its psycho-educational approach is designed to kickstart the mental processes responsible for attention, perception, reading and math, and the employment of both fine and gross motor skills by engaging children in playful activities that instruct and train the brain. It also aims to maximize children's early learning experiences through the stimulation of multiple sensory channels (e.g., auditory, visual, motor). As such, PlayWisely represents a unique and principled approach to educating children below the age of 3.

Method

Participants

Nineteen children took part in this pilot study (6 females and 13 males). All of them were recruited from the Kinderville Early Learning Centre in Ottawa, ON. After an initial round of pre-testing, the children were randomly assigned to either an immediate intervention group (4 females and 7 males, M = 18.70 months of age) or a wait-list control group (3 females and 5 males, M = 20.85 months of age).

Intervention

PlayWisely involves both cognitive and physical/motor skills components. For the cognitive component, a flashcard system is used that is specifically designed to stimulate the development of key cognitive processing abilities and knowledge structures (as opposed to the simple drilling of memory-based factual information). The cognitive skills targeted by different sets of cards in this system include visual perception and attention (e.g., signal detection, acuity, visual-spatial tracking, and speed of processing), discrimination of key physical-based attributes (i.e., shape, colour, shading, size, and part-whole relationships), concept acquisition (i.e., classification and object recognition), and multi-sensory information integration (i.e., across the auditory and visual modalities). They are also designed to develop foundational cognitive abilities associated with future numerical and literacy skills (e.g., subitizing, counting, digit and alphabet learning, phonemic awareness, and word recognition), hence, helping to ensure high levels of school readiness. Importantly, the skills targeted by each of the card sets in this system are highly specific and are trained in a developmentally appropriate manner that depends on the age of the child. Moreover, the kinds of visual stimuli used by the card system are chosen to be inherently interesting to young children (e.g., bugs, soccer balls, pig faces, etc.; as well, note that the cards are 10 inches by 10 inches in size), the events on each card are always simultaneously verbalized by the instructor in a rhythmic and encouraging manner, and there is a built-in regularity of events across sessions that allows children to learn to build-up expectations of what is coming next (i.e., the “rules of the game”).

The physical/motor component of each PlayWisely session involves training on 3-5 exercises within one of eight targeted skill domains: catching and throwing, balance, kicking, strength, fine motor, body translations, striking, and body rotations. The corresponding exercise sets are designed to enhance children’s awareness of how their body interacts with the physical environment and to enable a high level of brain-body confidence. These exercises are also performed in a developmentally appropriate manner that depends on the age of the child (e.g., there are Babies I, Babies II, Toddlers I, and Toddlers II versions of each exercise for children between the ages of ½ - 3 years).

Test Instrument

For the cognitive and motor ability testing, the Merrill-Palmer-Revised (M-P-R) Scales of Development (Roid & Sampers, 2004) was used. The M-P-R is an individually administered, standardized assessment instrument used to measure developmental skills in children aged 1 month through 6½ years. The gross motor and cognitive scales (the latter of which measures abilities related to reasoning, memory, both fine motor and visual motor skills, receptive language, and speed of information processing) are administered by an examiner and provide a developmental-index-based, age-equivalence score. The assessment is based on children's responses to toy-based activities prepared by the examiner and administration time is 45-60 minutes.

Procedure

In the immediate intervention condition, children participated in PlayWisely sessions once a week for 16 weeks (administered by the first author who is a trained PlayWisely instructor). The PlayWisely system itself involves eight different versions of the cards sets and eight different sets of physical/motor exercises. Hence, 16 weeks represents two full cycles through the PlayWisely materials. The wait-list control condition involved waiting until the immediate intervention group had finished with PlayWisely before starting their 16 weeks of PlayWisely sessions.

Participants were tested on the M-P-R Scales of Development at the very start (pre-test), again after the first 16 weeks of PlayWisely sessions with the immediate intervention group (post-test), and finally once more after the second 16 weeks of PlayWisely sessions with the wait-list group (follow-up), which resulted in a study length of 10 months after taking the additional time to perform each round of testing into account. All the tests of cognitive and gross motor abilities were administered by an examiner who was blind to the respective conditions assigned to each of the children being tested.

Results

One participant was dropped from each of the immediate intervention and wait-list control groups (a male and female, respectively) because they were unavailable for the final round of follow-up testing (leaving n = 10 and n = 7 in each group, respectively). The dependent measures for each of the following analyses were the age-equivalent scores (in months) derived from the M-P-R cognitive and gross motor ability scales. Each set of scores was subjected to a 3 (Test Period: pre-test, post-test, and follow-up) × 2 (Group: immediate intervention, wait-list control) ANOVA in SPSS with test period and group as repeated measures and independent groups factors, respectively.

For the statistical tests involving the M-P-R cognitive scale, the main effect of test period was significant ($F[2,30] = 180.02$, $p < .001$, $\eta_p^2 = .923$) but the main effect of group was not ($F[1,15] = 0.01$, $p < .928$, $\eta_p^2 = .001$). Although the overall Test Period × Group interaction was not significant ($F[2,30] = 2.03$, $p < .152$, $\eta_p^2 = .119$), a planned orthogonal contrast analysis did reveal that the linear trend of this interaction was significant at a marginal or one-tailed level of significance ($F[1,15] = 3.08$, $p < .100$, $\eta_p^2 = .170$; with the overall increase from pre-test to follow-up in mean age-equivalent cognitive ability score being 12.28 months for the intermediate intervention group but 9.97 for the wait-list control group).

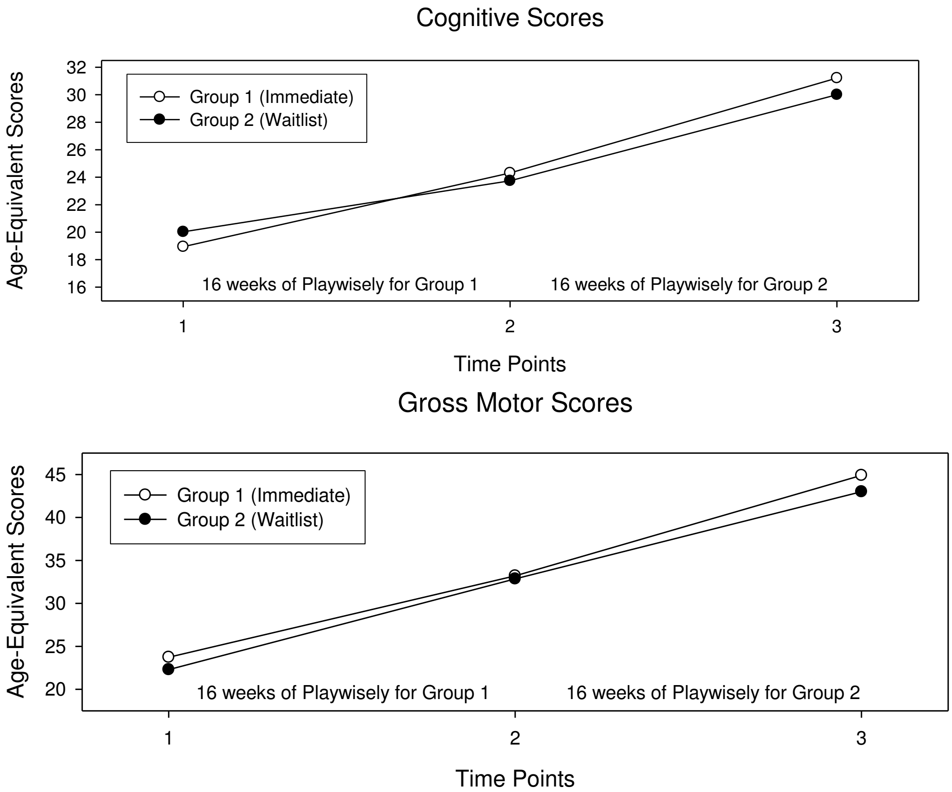


Fig. 1. Age equivalent cognitive and gross motor scores at pre-test, post-test, and follow-up.

For the statistical tests involving gross motor ability, the main effect of test period was significant ($F[2,30] = 128.76$, $p < .001$, $\eta_p^2 = .896$) but the main effect of group was not ($F[1,15] = 0.05$, $p < .827$, $\eta_p^2 = .003$). Here, the overall Test Period × Group interaction was also not significant ($F[2,30] = 0.23$, $p < .731$, $\eta_p^2 = .015$) and the planned orthogonal contrast analysis did not reveal any significant trends associated with this interaction (with the overall increase in mean age-equivalent gross motor ability score being 21.17 months for the intermediate intervention group and 19.71 for the wait-list control group).

Discussion

The results of this pilot study provide some preliminary evidence for a larger increase in age-equivalent cognitive ability scores over the testing period for the children who were administered PlayWisely immediately as opposed to 5 months later. Indeed, the increase for the immediate intervention group of 12.28 months was larger than would be expected given the actual time difference of 10 months between the pre- and follow-up tests (where presumably the wait-list children would also have ended up showing a comparable enhancement if the testing period could have been extended for a second, additional follow-up period). Importantly, the fact that this increase occurred for the M-P-R cognitive scale indicates that the effect of immediately receiving PlayWisely was to induce a general enhancement in cognitive ability.

On the other hand, these findings did not extend to age-equivalent gross motor ability scores. One reason for this is likely that the M-P-R gross motor scale did not perform very well (to the extent of suggesting age-equivalent increases that were actually twice as large as the testing period itself). Therefore, some consideration should be given to replacing this scale in future studies.

References

Kern, J.K., Garver, C. R., Mehta, J. A., Motes, M.A., Hannan, P. A., Bakken, L. E., Abraham, J., Viduad, A. M., & Daoud, Y. (2013). Prospective, blinded, exploratory evaluation of the PlayWisely program in children with autism spectrum disorder. *Yale Journal of Biology and Medicine*, 86, 157-167.

Roid, G.H., & Sampers, J. L. (2004). *Merrill-Palmer-Revised Scales of Development.* Wood Dale, IL: Stoelting Company.

PHASE TRANSITIONS IN BINARY CATEGORIZATION: EVIDENCE FOR DUAL-SYSTEM DECISION MAKING

Ihor Lubashevsky, Ian Wilson, Rio Suzuki, and Yuki Kato
University of Aizu, Ikki-machi, Aizu-Wakamatsu, Fukushima, 965-8580, Japan
i-lubash@u-aizu.ac.jp, wilson@u-aizu.ac.jp

Abstract

We report experiment results on binary categorization of (i) gray color, (ii) speech sounds, and (iii) number discrimination. Data analysis is based on constructing psychometric functions and focusing on asymptotics. We discuss the transitions between two types of subjects' response to stimuli presented for two-category classification, e.g., visualized shade of gray into "light-gray" or "dark-gray." Response types are (i) the conscious choice of non-dominant category, described by the deep tails of psychometric function, and (ii) subjects' physical errors in recording decisions in cases where the category choice is obvious. Explanation of results is based on the concept of dual-system decision making. When the choice is obvious, System 1 (fast and automatic) determines subjects' actions, with higher probability of physical errors than when subjects' decision-making is based on slow, deliberate analysis (System 2). Results provide possible evidence for hotly debated dual-system theories of cognitive phenomena.

Introduction

In the present work we report the results of our experiments on binary categorization of stimuli from two different sensory modalities (auditory and visual) and objects (numbers) whose perception is pure mental. Conducting these experiments, we pursued two goals:

- To find evidence for the universal mechanism of decision making in categorization which assumes that
    - the corresponding sense organ just converts an external stimulus into a neurophysical signal bearing the information about the stimulus proximity to the analyzed categories encoded in some general manner;
    - the brain processes this signal in a way independent of particular details characterizing the involved sensory modality.
- To verify whether it is possible to discriminate between situations when the decision-making in categorization is governed solely by physiological mechanisms and when deliberate analysis contributes substantially to categorization.

The reasons for posing these issues are as follows. First, the data accumulated in fMRI, EEG, and neuropsychological investigations enabled Walsh and Bueti (Walsh 2003, Bueti and Walsh 2009) put forward a new paradigm about human judgment and evaluation of external stimuli called ATOM ("A Theory Of Magnitude"). The ATOM supposes that various dimensions of magnitude information are encoded by "common neural metrics" in the parietal cortex, which explains the emergence of common neurocognitive mechanisms governing human perception of various physical stimuli. Hayes et al. (2014) generalized the ATOM by extending its scope onto memory, reasoning, and categorization. traditionally treated as separate components of human cognition. The exemplar-based account of the relationship between categorization and recognition (Nosofsky et al. 2012) is also rather close to this paradigm.

Second, as demonstrated (Baird and Noma 1975, Noma and Baird 1975) human perception of physical stimuli based on our sense organs and the mental evaluation of abstract objects like numbers are similar in the basic properties.

Third, nowadays the dual-processing account of human behavior is widely used in cognitive psychology. It holds that there are two distinct processing systems available for cognitive tasks. System 1 is fast, automatic and non-conscious, System 2 is slow, controlled and conscious. Whether the two systems do exist at the level of neurological processes or they admit the interpretation as individual subsystems with own properties is a subject of on-going debates, for arguments for and against a reader may be refereed to Rustichini (2008), Evans (2008, 2011), Kahneman (2011), Barrouillet (2011). Moreover, nowadays the idea about the cumulative contribution of two different mechanisms – fast guesses and slow controlled decisions – to the speed-accuracy tradeoff (Ollman 1966) has become popular; for a review see, e.g., Heitz (2014).

The purpose of our experiments was to accumulate enough statistical data to analyze the *asymptotics* of the corresponding psychometric functions accompanied with the dependence of the mean decision time on the uncertainty in category choice. As we demonstrated previously (Lubashevsky and Watanabe 2016, Namae et al. 2017), the asymptotics of psychometric function bears the information enabling one to discriminate between plausible mechanisms governing the categorization process in a clear way.

Gray color categorization

Each trial of color categorization was implemented as follows. A random integer $I \in [0, 255]$ is generated and some area on PC monitor is filled with the gray color $G(I) := RGB(I, I, I)$. Then a subject has to classify the visualized gray shade $G(I)$ according to his/her perception into two possible categories, "light gray" and "dark gray." A made choice is recorded via pressing one of two joystick buttons. Then a mosaic pattern of various shades of gray is visualized for 500 ms to depress a possible interference between color perception in successive trials that can be caused by human iconic memory. After that a new number I is generated and the next trial starts. The moment when a subject presses the button are also recorded, which gives the decision time in the current trial.

The experiments were set up as follows. Four subjects, two female and two male students of age 21–22 were involved in these experiments. The experiments spanned 5 successive days, for each subject the total number of data records was 2,000 data-points per day and finally 10,000 for 5 days. One day set comprised four blocks of 15 min experiments separated by 3 min rest. For each subject the total data-set aggregates all the records collected during 5 days. No special instructions were given to the subjects about the necessity to make decision in selecting categories as quickly as possible.

Figure 1 illustrates the obtained results which allow us to draw the following conclusions.

1. In the log-normal scales the asymptotics of psychometric functions for both the categories can be approximated by a linear dependence on the number I, which is the characteristic feature of potential mechanism (Lubashevsky and Watanabe 2016). It means that the decision-making may be regarded as a probabilistic event of finding a particle in an equilibrium state described by some potential. The region of this asymptotic behavior matches the most pronounced uncertainty in selecting the categories.
2. As the analyzed gray shade (the number I) penetrates deeper in the region of rare events (rare choice of inappropriate category) this linear asymptotics is replaced by the probability of choosing the inappropriate category that does not decrease or even can increase with the further change in the number. The transition between these modes of the choice probability behavior is well pronounced, which is noted by arrows in Fig. 1.
3. The increase of the mean decision time in the region of choice uncertainty typically exceeds 1 s, which is substantially longer than the upper boundary of human response delay controlled by pure physiologically processes. It argues for a significant contribution of mental processes to the categorization in the region of its essential uncertainty.

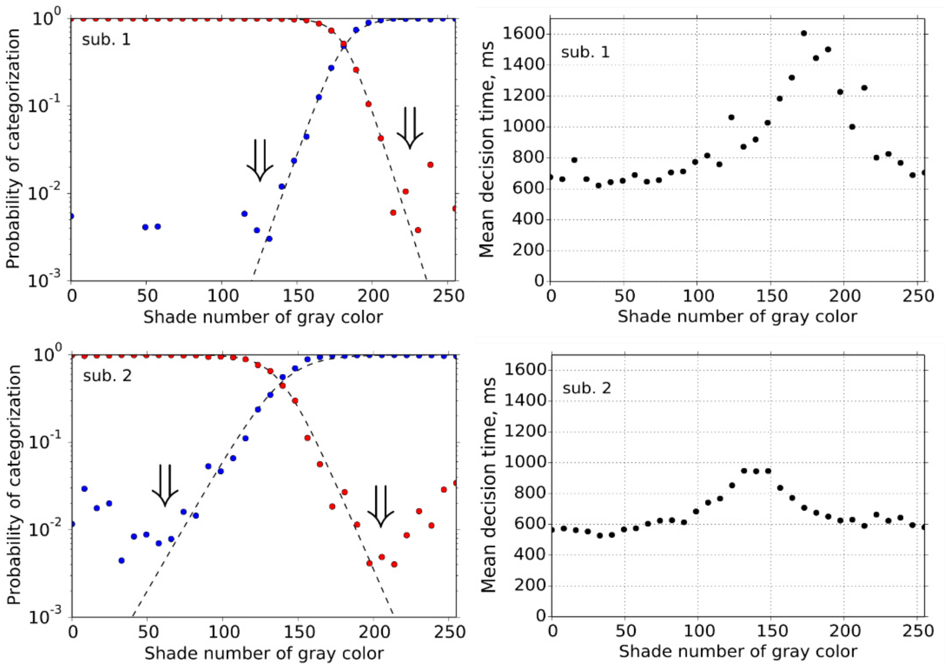


Fig. 1. *Left column:* Psychometric functions of gray color categorization, data points corresponding to the “light-gray” and “dark-gray” classes are shown in blue and red, respectively. Dashed lines represent fitting functions of logistic form. *Right column:* Mean decision time in gray color categorization depending on the shade number of gray.

Vowel sound categorization

Using the "Create Sound from Vowel Editor…" menu item in Praat acoustic analysis software (Boersma & Weenink, 2018), we synthesized a total of 68 sound files. Each sound was a 100-millisecond vowel with a different second-formant (F2) frequency, which ranged from 810 Hz ([o]) to 2,150 Hz ([e]) in 20 Hz steps. The first-formant (F1) was kept constant at 400 Hz for all sounds. So, the first sound's (F1, F2) values in Hz were (400, 810); the second sound's values were (400, 830), the third sound's values were (400, 850), and so on up to the 68th sound, which had values (400, 2150). Thus, the 68 sounds varied along a single dimension – that of F2. The pitch of all sounds was kept constant at 140 Hz–a pitch within the normal speaking range of a human voice. The 68 unique audio files were randomly ordered to make a "cycle", and this was done 12 times to make 12 different cycles. A randomly ordered group of 12 cycles was considered 1 trial and contained 68 x 12 = 816 vowel sounds. Each subject completed 10 trials (each with a random order of cycles) with a short break between each one, for a total of 8,160 vowel sounds.

Two healthy subjects with no reported hearing problems participated in this research. They were both Japanese 4th-year undergraduate students who were in their early twenties. Subject 1 was female and Subject 2 was male. E-prime 2.0 software (*Psychology Software Tools*) and a 5-button "Chronos" multifunctional response and stimulus device were used for recording subject actions. The Chronos device has millisecond accuracy and consistent sound output latencies across machines. Subjects listened to the stimuli through JVC headphones and were instructed to react as quickly as possible to identify the stimuli that they heard.

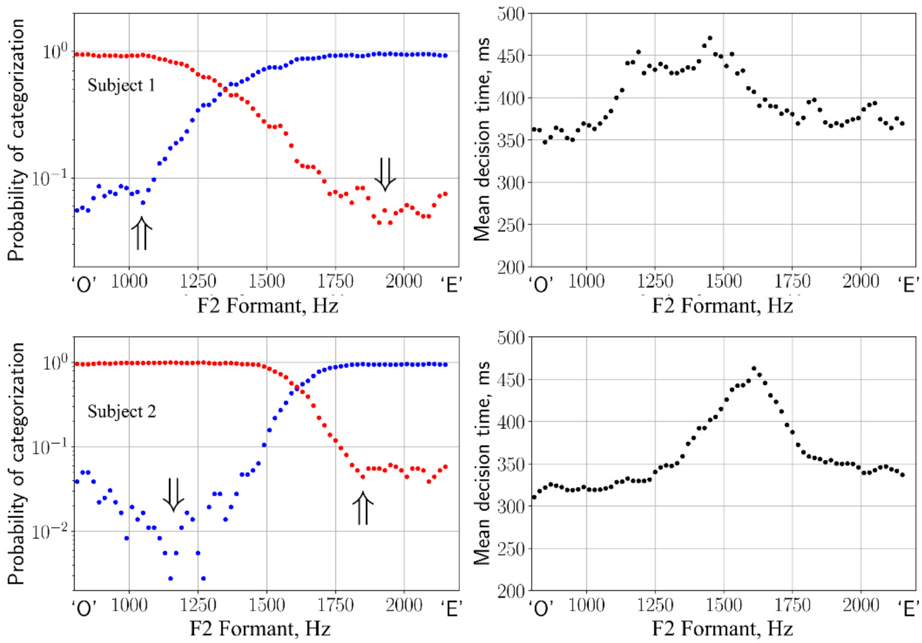


Fig. 2. *Left column:* Psychometric functions of vowel sound categorization, data points corresponding to [o] and [e] classes are shown in red and blue, respectively. *Right column:* Mean decision time in vowel sound categorization depending on the F2 Formant.

Figure 2 exhibits the obtained results which also argue for conclusions 1 and 2 stated in the previous section for the color categorization. The main difference is that the mean decision time in its maximum attains the upper boundary of the response delay time usually attributed to the information processing by neural networks. However, the similarity of the asymptotic behavior of psychometric functions in both the cases allows us to hypothesize that mental processes play a significant role also in the sound categorization in the region of its substantial uncertainty. The reaction of subjects instructed to respond as quickly as possible just corresponds to the minimal delay of the conscious response coinciding with the maximum of physiological delay such that no gap in the human response delay time appears.

Categorization of numbers

The idea of the experimental setup of studying the binary categorization of numbers is similar to that of color categorization. An integer I from the interval [1, 256] is randomly selected and visualized. A subject has to select or reject a visualized integer, which is recorded via pressing the corresponding button of joystick. The conditions of experiments stimulated subjects to select integers near the right boundary of the interval [1, 256] and to reject integers near its left boundary with certainty. The uncertainty of this choice becomes significant for intermediate integers and the width of the corresponding region is estimated as 50. Four subjects – male students of about 20 years old – were involved in these experiments. The experiments were done for 4 sets with about 2,000 data points per set; the total amount of recorded data for each participant is about 8,000 data-points.

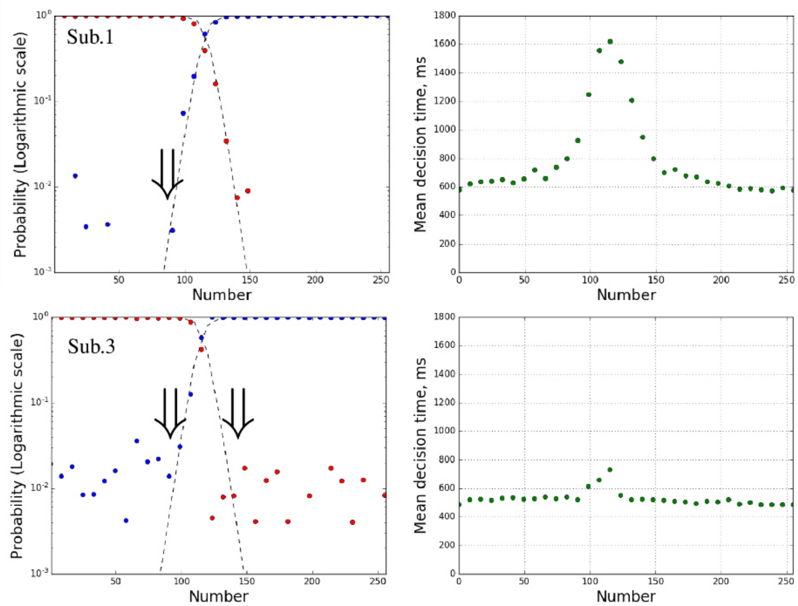


Fig. 3. *Left column:* Psychometric functions of number categorization, data points corresponding to “rejecting” and “accepting” a given number are shown in red and blue, respectively. Dashed lines represent fitting functions of logistic form. *Right column:* Mean decision time in number categorization (based on Nihei, 2018).

Figure 3 illustrates the obtained results. The categorization of numbers under the given conditions may be treated as some mixture of the color and sound categorization in property.

Conclusion

The presented results enable us to posit the following.

- The choice between the auditory and visual categories as well as abstract categories in number comparison is governed by a universal central mechanism of the potential type. It is reflected in the linear asymptotics of psychometric functions in log-normal scales, which corresponds to significant uncertainty in categorization.
- There is a sharp transition between the linear asymptotic behavior of psychometric functions and their behavior when the choice of appropriate category is obvious. We relate the linear asymptotic behavior to conscious choice governed by slow System 2; it is characterized by the regular growth of decision time as the choice uncertainty increases. The choice of appropriate category when it is obvious seems to be governed by automatic, fast System 1. It explains the relative increase in the probability of choosing the inappropriate category by human errors in motor behavior when the conscious control is depressed.
- When the choice uncertainty is high, mental processes seem to affect substantially the decision time and its maximum can be used to quantify the relative contribution of conscious and unconscious processes as a whole. In particular, the shorter the maximal decision time, the higher the contribution of unconscious component, the higher the probability of motor errors in the case of System 1 control.

Acknowledgements

The authors wish to acknowledge the University of Aizu for Competitive Research Funding that was used in carrying out this research.

References

Baird, J. C. and Noma, E. (1975). Psychophysical study of numbers: I. Generation of numerieal responses, *Psychological Research* **37**(4), 281–297.

Barrouillet, P. (2011). Dual-process theories and cognitive development: Advances and challenges, *Developmental Review* **31**(2–3), 79–85. Special Issue: Dual-Process Theories of Cognitive Development.

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.40, retrieved 11 May 2018 from http://www.praat.org/

Bueti, D. and Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes, *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**(1525), 1831–1840.

Evans, J. S. (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.

Hayes, B. K., Heit, E. and Rotello, C. M. (2014). Memory, reasoning, and categorization: parallels and common mechanisms, *Frontiers in Psychology* 5, Article 529 (9 pages).

Heitz, R. P. (2014). The speed-accuracy tradeoff: history, physiology, methodology, and behavior, *Frontiers in Nuroscience*, **8**, Article 150 (19 pages).

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

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

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

Nihei, K. (2018). Statistical Properties of Number Classification: Asymptotics of Psychometric Function and Mean Decision Time, Thesis, University of Aizu, supervised by Lubashevsky I.

Noma, E. and Baird, J. C. (1975). Psychophysical study of numbers: II. Theoretical models of number generation, *Psychological Research* **38**(1), 81–95.

Nosofsky, R. M., Little, D. R. and James, T. W. (2012). Activation in the neural network responsible for categorization and recognition reflects parameter changes, *Proceedings of the National Academy of Sciences of the United States of America* **109**(1), 333–338.

Ollman, R. T. (1966). Fast guesses in choice reaction time, *Psychonomic Science* **6**(4), 155–156.

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

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity, *Trends in Cognitive Sciences* **7**(11), 483–488.

BAYESIAN HIERARCHICAL MODELLING OF GOAL SIDE SELECTION IN A SOCCER PENALTY KICKING TASK

Geoffrey R. Patching and Mauro R. Pereira
Department of Psychology, Lund University, Sweden
Geoffrey.Patching@psy.lu.se

Abstract

The present study investigates goal side selection in soccer penalty kicking. Given the task of choosing which side of a soccer goal to best score, participants viewed realistic images of a soccer goal and goalkeeper. The goalkeeper's position was systematically displaced along the goal line, and the lateral position of the goalmouth was systematically displaced in each image – to simulate changes in the viewing position of the kicker. Hierarchical Bayesian analysis is used to examine participant's weighting of the goalkeeper's position and kicker's viewing position on binary goal side selection. Overall, participants tended to choose the left over right goal side, but this depended on the goalkeeper's position relative to the centre of the goal and, to lesser extent, lateral position of the goalmouth relative to the participant's body midline. These findings are considered in relation to the line bisection task as regularly used in neuropsychology to assess visual neglect.

Recently a number of studies have emerged (Masters, et al., 2007; Noël, et al., 2015a, 2015b) which suggest that penalty kickers tend to place the ball to the goal side with greater area when the goalkeeper stands marginally to the left or right of the true centre of the goal line. This so called off-centre effect has been found in analysis of high level soccer competitions (Masters, et al., 2007), penalty kicks taken against images of a goalkeeper and a soccer goal (Masters, et al., 2007), and in trials on a soccer pitch (Noël et al., 2015a; 2015b), even when participants report that the goalkeeper is standing centrally (Masters, et al., 2007; Noël et al., 2015a, 2015b).

Here, an aim is to contribute to current understanding of relations between goal side selection and goalkeeper position by systematic displacement of the goalkeeper's position along the goal line and by simultaneous displacement of the goalmouth position relative to the participant's body midline – to simulate changes in the viewing position of the kicker. To date there is a dearth of research examining the influence of the goalkeeper's position and simultaneously initial viewing position of kicker before taking the shot. Yet, most kickers approach the ball at an angle which in turn may determine (or at least play a role) in goal side selection. As a result, goal side selection in soccer may not only be influenced by the goalkeeper's position but may also depend on the initial viewing position of the kicker.

Given changes in the goalkeeper's position alone, Masters et al., (2007) found that participants kicked a ball to the larger goal side area on about 60% of trials for differences in area of just 0.5%, which increased to about 95% of trials for differences in area of 5%: where the difference in area was calculated as the difference between the two goal areas either side of the goalkeeper's midline as a percentage of total goal mouth area (Δarea / area * 100). So, it appears that the proportion of left goal side choices rises monotonically from 0 to 1, as the position of the goalkeeper is systematically displaced along the goal line from left to right of the goal midline, describing a generalized psychometric function.

One approach employed in modelling psychometric functions is to fit individual curves to each participant's data and then perform statistical analysis on the extracted parameters. An alternative to this two stage approach is to use Generalized Linear Mixed Effects Modelling (GLMM; Moscatelli, et al., 2012). GLMM is advantageous to the classic two stage approach because it permits modelling all of the data and all of the parameters in one step rather than two. However, when using maximum likelihood methods, such models all too often fail to converge due to the large parameter space to be searched. Markov chain Monte Carlo (MCMC) methods, as used in Bayesian data analysis, overcome this issue and provide estimates of all credible values of all parameters of interest.

After a hierarchical Bayesian model has been fitted the estimated parameters can be investigated in many ways. The parameters can be examined at the participant level or the group (hyper parameter) level, depending on where interest lies. Because the model is fully Bayesian, estimates for all of the parameters are readily available after the model has been fitted, including reliability measures in the form of credible intervals, here presented in terms of the 95% highest density interval (HDI) as advocated by Kruschke (2015). In the present study, GLMM revealed the same point estimates of the parameters as revealed by Bayesian hierarchical modelling. But Bayesian hierarchical modelling is more informative than GLMM, and so the results of the GLMM analysis is not reported.

Method

Participants

40 participants took part in the experiment of whom 11 were women and 29 men, aged between 21-36 years (mean 29 years), recruited from Lund University's student population. All claimed to be right handed. All reported normal or corrected-to-normal vision, and all but 3 participants claimed to be right-footed. None of the participants played soccer on a regular basis.

Stimuli / design

The stimuli consisted of 16 images presented on a standard computer monitor (Fujitso Lifebook Series 5). Each image was 185 x 156 mm in size. The goal mouth dimensions depicted in the images was 140 x 49 mm (0.0069 m^2), which is 0.04% of the total area of original sized goals used in association football (2.44 x 7.32 m [17.86 m^2]). The goalkeeper's height was 40 mm [approximately 2% of Manuel Neuer's real height (1.93m)], and the distance between the goal line and the penalty spot (where the ball was shown) was scaled to 0.3% (0.03m) of real playing distance (11m).

The goal keeper was positioned at 7 different locations relative to the centre of the goal, from -3.44 mm (left) to +3.44 mm (right) in 7 steps of 1.13 mm. In addition, the whole goal mouth was presented at 7 different locations relative to the centre of each image / computer monitor, from -3.44 mm (left) to +3.44 mm (right) in

7 steps of 1.13 mm. After Hellström (1979) the 7 goalkeeper and 7 goalmouth positions were combined semi-factorially to form a ‘diamond’ shape about centre, to create 16 unique images characterized by different combinations of goalkeeper and goalmouth displacements. The physical differences between the two image displacements ranged from -3.14 to +3.14 mm in 4 steps of 1.13 mm across one diagonal, and the mean average of the two displacements ranged from -1.70 mm to +1.70 in 4 steps of 1.13 mm across the opposing diagonal.

Procedure.

The experiment consisted of 2 sections (practice and experimental), with no break between the 2 sections. The first 32 trials, in which each stimulus was presented twice in pseudorandom order, were deemed practice trials. Following the practice trials, participants completed a further 256 experimental trials in which all the stimuli were presented 16 times in pseudorandomised cycles of 64 trials.

Twenty participants were instructed to indicate right goal side by pressing the up arrow key with the forefinger of their right hand and left goal side by pressing the down arrow key with the index finger of their right hand. The other 20 participants were instructed to indicate right goal side by pressing the down arrow key with the index finger of their right hand and left goal side by pressing the up arrow key with the forefinger of their right hand. Participants were seated comfortably, aligned centrally to the computer monitor at arm’s length (~57 cm).

From the kicker’s perspective, participants were instructed to decide as quickly as possible the best side of the goal (left or right) to place the ball to score a goal. On each trial, each image was presented until the participant made a goal side selection either by pressing the up arrow key or down arrow key. Reaction time (RT) was measured from stimulus onset until the participant made a response. The inter-trial-interval was set at a random duration from 1000-3000 ms. On the average, participants took 30 minutes to complete the experiment. For brevity, the RT data are not presented in the present paper. Analysis of the RT data revealed a similar pattern of results as revealed by analysis of participants’ binary goal side selections.

Data analysis

One participant was removed from all further data analysis, because they consistently chose the left goal side over all trials. The remaining (N = 39) participants’ relative weighting of the goalkeeper’s position and kicker’s position in goal side selection was assessed by logistic regression of each participant’s binary responses on the goalkeeper’s position and kicker’s position. For comparability with earlier research (Masters et al., 2007), the goalkeeper’s position is expressed in terms of the difference between the left and right goal areas either side of the goalkeeper’s midline as a percentage of total goal mouth area [(left area - right area) / total area * 100]; hereafter, referred to as the goalkeeper’s position. Likewise, the kicker’s position is expressed in terms of the difference between the left and right goal areas either side of the veridical centre of the kicker’s egocentric viewing position as a percentage of total goal mouth area; hereafter, referred to as the kicker’s position.

The relative influence of the goalkeeper’s position and kicker’s position on goal side selection is then captured by way of a simple linear equation in which the goalkeeper's position (*keeper*) and kicker's position (*kicker*) are multiplied by a weighting (β_1 and β_2), the products summed and a constant (β_0) added. Let GS_{12} equal goal side selection in terms of logit $P = \log_e[p / (1 - p)]$: the natural logarithm of the proportion of left goal side responses over the proportion of right goal side responses then, $GS_{12} = logistic(\beta_0 + \beta_1 keeper + \beta_2 kicker)$, where $logistic\,(x) = 1/(1 + \exp(-x))$. Here, the coefficients β_1 and β_2 capture the relative weighting placed on the goalkeeper's position and kicker's position in goal side selection. Initial examination of the data did not reveal any systematic effects of manipulating response assignment, so this variable is not included in any of the models reported in the present paper.

In terms of hierarchical Bayesian analysis, GS_{12} is described as distributed around the standardized lateral position of the goalkeeper (*Zkeeper*) and kicker's position (*Zkicker*) plus a constant, β_0. The goalkeepers position and kicker's position are mapped to a probability value via the logistic function, and so predicted binary goal side selection is Bernoulli distributed around this probability value.

$$\mu = \text{logistic}\,(\beta_0 + \beta_1 Zkeeper + \beta_2 Zkicker)$$
$$y \sim \text{Bernoulli}(\mu)$$

For each participant the parameters β_0, β_1, and β_2 were estimated following normal distribution functions parameterized by mean μ and standard deviation σ. Over all participants, group level hyper priors for these parameters followed normal distributions where μ_μ followed a vague normal distribution with mean zero and standard deviation of 10000, and σ_σ followed a gamma distribution with mean one and standard deviation of 100. Varying the standard deviation of the priors had no noticeable effect on the subsequent results obtained.

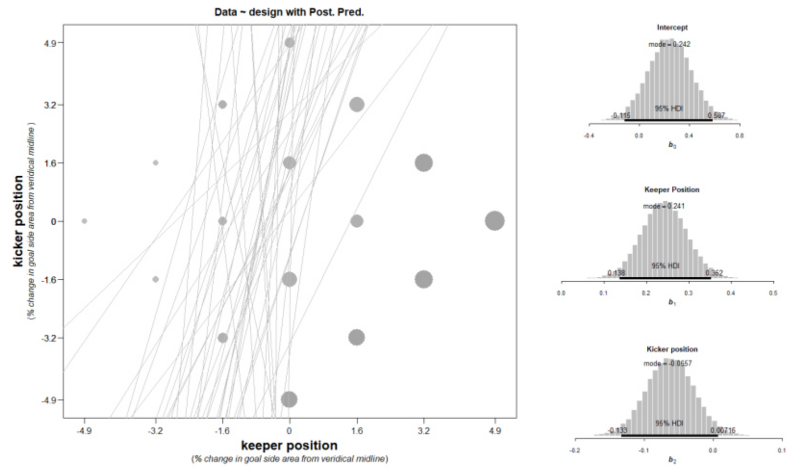


Figure 1. Left panel shows the data following the design of the experiment. The larger the dots the greater the proportion of left goal side selections. The grey lines superimposed on the design show plausible 50% level contours. Right panel shows group level estimates for the unstandardized coefficients b_0, b_1 and b_2, as found by hierarchical logistic regression of each participant's binary choice responses on the goalkeeper's position and kicker's position.

The joint effect of the goalkeeper's position and kicker's position is then examined in terms of the 4 average positions of the goalkeeper and kicker depicted in the 16 images (Figure 1 lower left to upper right diagonal). The joint mean position of the goalkeeper and kicker for each stimulus image were -2.4, -0.8, 0.8 and 2.4, as calculated in terms of average percentage left goal area and corrected for rounding error. Participants' binary responses were then regressed on the 4 average positions of goalkeeper and kicker using Bayesian hierarchical modelling in exactly the same manner as described above, with the exception that there was only one predictor in this case.

All analyses were run with R (R Core Team, 2018) and JAGS (Plummer, 2003). All posterior distributions are based on 500 adaptation steps, 1000 burn-in steps, and 45,000 saved steps with no thinning, using 3 chains, which showed little autocorrelation. Good convergence obtained between the multiple chains as tested with the R statistic (Gelman, Carlin, Stern, & Rubin, 2004), which was found to be less than 1.1 in every case. The resulting MCMC samples of the posterior distributions are therefore highly representative of the underlying distributions.

Results

Figure 1 shows the results of predicting goal side selection from the goalkeeper's position and kicker's position. After Kruschke (2015), the left panel shows the study design with the dots scaled relative to the overall proportion of left goal side selections, *p*(left goal side selection)[2]. Superimposed on the data are 40 plausible level contours at which *P*(left goal side selection) = 50%. The 50% level contour is the set of x_1 = keeper position and x_2 = kicker position values for which $\mu = 0.5$, given by $x_2 = (-\beta_0 \,/\, \beta_2) + (-\beta_1 \,/\, \beta_2)_{x1}$. The spread of the level contours provides an indication of the uncertainty in the parameter estimate.

Tracing a line perpendicular to the level contour signposts the direction in which probability changes the fastest. In the present case, the angle of the contours suggests that the probability of left goal side selection increases steadily as the goalkeeper's position is moved from left to right along the goal line. Conversely the probability of left goal side selection decreased only a little as the kicker's viewing position was moved from left to right of veridical goal centre - the unstandardized regression coefficient (b_2) on kicker position has a small modal value of just -0.07 and its 95% HDI [-0.13, 0.01] crosses zero.

Figure 2 shows the results of predicting goal side selection from the 4 joint average positions of the goalkeeper and kicker. The left panel of Figure 2 shows the data plotted as points that fall at 0 and 1 on the *y*-axis (arbitrarily coded, left goal side selection = 1, right goal side selection = 0). The grey diagonal lines superimposed on the data are logistic curves that have plausible parameter values as sampled from the posterior MCMC chain.

The spread of the logistic curves indicates uncertainty in the parameter estimate. The 50% probability threshold, *P*(left goal-side selection) = 50%, is marked by dotted lines that drop down from the logistic curve to the *x*-axis, as given by $x = (-\beta_0 \,/\, \beta_1)$. The left panel of Figure 2 shows the marginal posterior MCMC estimates of the group level, hyper, parameters. In particular, the unstandardized slope coefficient b_1 has a mode of 0.15, and a 95% HDI [0.07, 0.24] that does not cross zero.

Discussion

Hierarchical Bayesian modelling provides for a powerful and flexible way to analyse psychophysical data. Within a hierarchical framework, individual participants are nested within a group allowing for assessment of individual differences and group level behaviour within a single model. Once the model has been fitted complete distributions of credible values are obtained for all parameters of interest, which can simply be sampled from the posterior MCMC distribution.

In the present study, hierarchical Bayesian analysis of goal side selection in a soccer penalty kicking task showed that participants' tended to choose the left over right goal side. On the basis that participants chose the goal side with the largest perceived area to best score a goal, this fits with studies of line bisection and the so-called "Landmark Task" (Harvey, Milner, & Roberts, 1995) in which healthy participants tend to overestimate the length of left as compared to right line segments (McCourt & Jewell, 1999). Moreover, the asymmetry found in goal side selection is reminiscent of systematic asymmetries found in comparison and discrimination of paired stimulus magnitudes; termed, time- and space-order effects (Patching et al., 2012). The present study shows that participants increasingly chose the left over right goal side as the goalkeeper's position was moved systematically from left to right along the goal line relative to goal midline. But, systematic changes in the egocentric viewing position of the kicker (i.e., participant) had far less effect. Given the predominance of left over right goal side responses and considerable individual differences in goal side selection, group level estimates of the parameters are somewhat uncertain. Further analysis of penalty kicks taken in professional soccer tournaments may increase the validity of this work. For those familiar with GLMM using R, the recent 'brms' package (Bürkner, 2017) allows users to fit a broad range of Bayesian linear and nonlinear models using accessible formula syntax without having to learn new computational code.

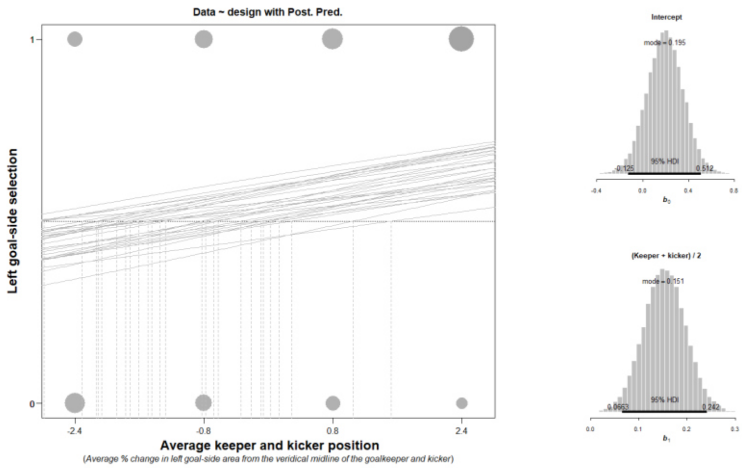


Figure 2: Left panel shows the data following the joint, average, positions of the goalkeeper and kicker. The points are scaled relative to the proportion of left and right goal side selections. The grey lines superimposed on the design show plausible logistic curves. The descending dotted lines point to average goalkeeper and kicker positions at which the probability of left goal side selection is 50%.

References

Bürkner, P-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*, 1.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2004). *Bayesian data analysis*. New York: Chapman and Hall.

Harvey, M., Milner, A. D., & Roberts, R. C. (1995). An investigation of hemispatial neglect using the Landmark Task. *Brain and Cognition, 27*, 59-78.

Hellström, Å. (1979). Time errors and differential sensation weighting. *Journal of Experimental Psychology: Human Perception and Performance,* 5, 460-477.

Kruschke, J. K. (2015). *Doing Bayesian Data Analysis. A tutorial with R, JAGS and STAN*. Elsevier, Amsterdam.

Masters, R., van der Kamp, J., & Jackson, R. (2007). Imperceptibly off-center goalkeepers influence penalty-kick direction in soccer. *Psychological Science*, *18*, 222-223.

McCourt, M. E., & Jewell, G. (1999). Visuospatial attention in line bisection: stimulus modulation of pseudoneglect. *Neuropsychologia, 37*, 843-855.

Moscatelli, A., Mezzetti, M., & Lacquaniti F. (2012). Modeling psychophysical data at the population-level: The generalized linear mixed model. *Journal of Vision, 12*: 26.

Noël, B., van der Kamp, J., & Memmert, D. (2015a). Implicit goalkeeper influences on goal side selection in representative penalty kicking tasks. *PLoS ONE*, 10.

Noël, B., van der Kamp, J., Weigelt, M., & Memmert, D. (2015b). Asymmetries in spatial perception are more prevalent under explicit than implicit attention. *Consciousness and Cognition, 34*, 10-15.

Patching, G. R., Englund, M. P., & Hellström, Å. (2012). Time- and space-order effects in timed discrimination of brightness and size of paired visual stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 915-940.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs Sampling, *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), March 20–22*, Vienna, Austria.

R Core Team (2018). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Weigelt, M., & Memmert, D., & Schack, T. (2012). Kick it like Ballack: The effects of goal-keeping gestures on goal-side selection in experienced soccer players and soccer novices. *Journal of Cognitive Psychology, 24*, 942-956.

Weigelt, M., & Memmert, D. (2012). Goal-side selection in soccer penalty kicking when viewing natural scenes. *Frontiers in Psychology, 3*, Article 312.

THE TARGET DETERMINES THE JOURNEY: LETTERS ELICIT SPECIFIC PROCESSING STRATEGIES IN CHARACTER-STRINGS

Andreas Schmitt[1,2] and Thomas Lachmann[1]
[1]*Center for Cognitive Science, University of Kaiserslautern, Germany*
[2]*Institute for Educational Science, University of Oldenburg, Germany*
andreas.schmitt@uni-oldenburg.de, lachmann@rhrk.uni-kl.de

Abstract

In five-character strings, letters are detected faster at the initial, central and final positions of the string together with an increase in reaction times from left to right, whereas for non-letter material the advantage for terminal positions vanishes, resulting in a symmetrical center-out increase in reaction times. To investigate the underlying processing strategies, in a target-detection task, participants had to decide whether or not a target, either a letter or a non-letter, was part of a five-character string, randomly mixed out of letters and non-letters. The typical patterns of reaction times were observed for either letters or non-letters targets. Since the five-character strings were always composed out of letters and non-letters, these results suggest, that the different processing strategies are not elicited by the string itself but rather automatically triggered by the target character. The results are interpreted in the context of modification and specialization of visual processing strategies during reading acquisition.

Automatization of reading-specific procedures makes letters special. Skilled readers perceive letters faster, but also differently than comparable non-letter shapes, even outside the context of reading, such as in letter recognition or classification tasks (Burgund, Schlaggar & Petersen, 2006). It was shown that the distinctive processing of letters is not simply an effect of higher stimulus familiarity, but reflects an Analytic Processing Preference for Letters Effect (APPLE; Lachmann, 2018) which was automatized during literacy acquisition (Lachmann and van Leeuwen, 2014). The APPLE occurs if the task requires alphabetic decoding or if such decoding is beneficial. For single letters the APPLE was demonstrated in several studies by distinctive symmetry effects (Lachmann & van Leeuwen, 2007; Pegado et al., 2014, 2011), flanker effects (Fernandes & Kolinsky, 2013; Lachmann & van Leeuwen, 2004, 2008; van Leeuwen & Lachmann, 2004) and by the absence of the Global Preference Effects (Poirel, Pineau & Mellet, 2008; Lachmann, Schmitt, Braet & van Leeuwen, 2014). Regarding the latter, Lachmann et al. (2014) demonstrated that in hierarchical compound figures reading-like conditions are crucial to elicit analytic processing for letters, whereas non-letters are still processed in a holistic manner. This letter versus non-letters distinction in hierarchical compound figures even persists after associating the non-letters with phonemes during several training sessions (Schmitt, van Leeuwen & Lachmann, 2017).

Although reading-like conditions seems to benefit analytic letter processing, the APPLE was mostly demonstrated for single letter processing. Hence, it remains unclear how analytic processing manifests, as a part of the acquired reading processing strategies, in the context of reading words or at least word-like material.

Since real-word-reading is also driven by lexical information and semantical context (Coltheart, Curtis, Atkins & Haller, 1993), these have to be excluded to detach the analytic component of reading strategies. Reading non-words (i.e. randomly composed character strings) complies these conditions.

Although the APPLE was explicitly demonstrated for single letters, various studies suggest that distinctive processing of letters and non-letters leads to different processing of character strings. In five-character strings, built-up of non-letters, there is an advantage for the central positions, i.e. characters are detected faster and more accurate compared to the terminal positions (Hammond & Green, 1982; Pitchford, Ledgeway & Masterson, 2008), resulting in a quadratic U-shaped function over the different positions for reaction times and error rates. This result was explained by sensory center-out scanning mechanisms, as a consequence of foveal perception of the central character, where visual acuity is greatest (Mason, 1982; Wagstaffe, Pitchford & Ledgeway, 2005) combined with a decrease of visual acuity from the foveal to the parafoveal retinal field (Wagstaffe et al., 2005).

However, for letter strings an advantage in reaction times and error rates was observed for the central position as well as for the first and final position. In addition, there is an overall increase in reaction times from left-to-right. Hence, for letter strings, the sensory explanation only holds for the advantage at the central position. The facilitated end-effects as well as the linear left-to-right trend were explained as resulting from literacy acquisition and acquired letter processing (Grainger, Tydgat & Issele, 2010; Stevens & Grainger, 2003) and reading habits, e.g. the serial decoding of letters in written words (Green et al., 1983; Hammond & Green, 1982; Ktori & Pitchford, 2008, 2009). Depending on language-specific morphological and orthographical constrains the linear left-to-right increase overweights the terminal position advantage. Thus, only the advantage for the initial and central position could be observed, but not for the final position (Farid & Grainger, 1996; Ktori & Pitchford, 2008, 2009).

In this study a target-detection task was used in which participants have to decide whether a pre-designated target character, either a letter or a non-letter, is part of a five-character string consisting of both, letters as well as non-letters. Thus, a distinction between letter and non-letter detection could only be caused by the material type of the target, not by the string that needs to be scanned.

As mentioned previously, the APPLE results from automatization during extensive and long-term reading experience. Since skilled reading mostly involves processing of complete words rather than single letters, specific processing should not be restricted to single letter processing. Analytical processing also includes processing of letter strings. The typical pattern in paradigms with letter strings, i.e. the left-to-right trend and the terminal position advantage, can be interpreted as a result of these analytical strategies. The analytical strategies are not a result of intentional choice, but of automatically elicitation if letter processing is required. Hence, the distinctive processing strategies for letters and non-letters are not caused by the type of characters in the string but by the type of the target character which should be processed. Thus, we would expect to observe the analytical strategies also in mixed character strings, i.e. randomly compound character strings of letters and non-letters, if these strategies are elicited by a pre-designated letter target.

In accordance to the typical results for letter-strings this analytic processing would include a linear left-to-right increase for target detection times as well as a terminal position advantage. Although the task and the character-strings do not differ, analytic processing should not be elicited by non-letter targets. Hence, if non-letters have to be detected the character-strings still should be processed holistically, i.e., a symmetric center-out scanning with larger detection times for the terminal than for the inner positions.

Method

Note: The following sections just give a summary of the method and results. Full information could be received in Schmitt and Lachmann (in progress); see also Schmitt, (2016).

Participants

Forty-one students (20-30 years, right handed) from the University of Kaiserslautern (Germany).

Material

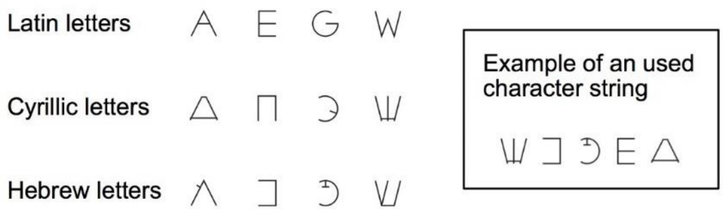


Fig. 1. Illustration of the characters. The characters consisted of four different Latin letters as letters and four different Hebrew and Cyrillic letters with similar complexity but unknown as letters for the participants and were thus used as non-letter characters.

Procedure

The experiment consisted of 480 trials spread over two sessions performed on one day with a short break in-between. In a target-detection task, participants had to decide whether or not a pre-designated target was part of the presented character string (see Fig. 1). Participants responded by pressing "s" with their left index finger or "l" with their right index finger on the keyboard. The assignment of keys and response alternatives (character is / is not part of the string) was counterbalanced between participants. Each character appeared as target in 40 trials. In half of these trials, the target was absent, i.e., the target character was not part of the character-string. The remaining 20 trials were counterbalanced in a way that the target was present for four times on each position of the five-character string. In sum, each kind of characters (Latin, Hebrew, and Cyrillic) was shown 16 times on each position. Trials began with the presentation of the target character in the center of the screen for 1000 ms, followed by a blank screen for 500 ms. The five-character string was shown centrally, until the participants responded. After a blank screen, shown for 1500 ms, the next trial began. Participants completed 30 training trials, providing feedback for 1000 ms about their individual reaction times and accuracy.

Results

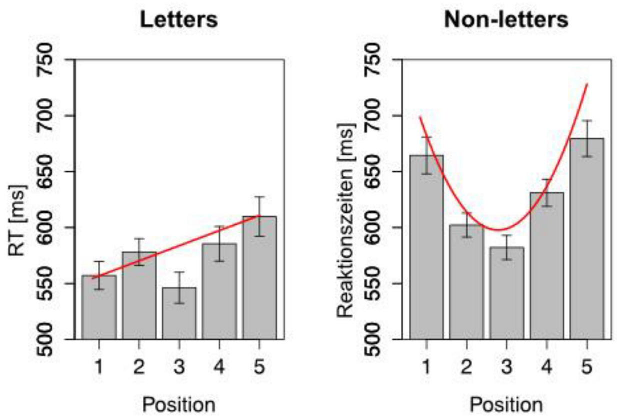


Fig. 2. Illustration of reaction times (RTs) and model fits.

Since preliminary analyses did not show any difference between Hebrew and Cyrillic letters, these characters were merged to the factor level *non-letters*. Mean reaction times (RTs) for correct responses within a range 150 ms and 2000 ms (95 trials excluded, 0.5 %) were analyzed in a repeated measures 2 x 5 ANOVA with the factors *Material* (letters vs. shapes) and *Position* (1-5 from left to right in the character-string).

Main effects were observed for *Material*, $F(1, 38) = 44.37$, $p < .001$, with faster responses for letters than for non-letters and *Position*, $F(4, 152) = 28.03$, $p < .001$, showing that responses are faster for the central position and become slower the further they are away from the central position. There was also a *Material x Position* interaction, $F(4, 152) = 15.31$, $p < .001$.

Further post-hoc analyses showed, that the interaction was caused by differences in performance of letters and shapes at the first position in the strings. For shapes, the mean RTs described a U-shaped position function with slow responses for the first position, fastest responses for the central position and again slow responses for the final position. For letters, there was no such U-shaped position function, resulting from fast RTs at the first position, which did not differ significant from RTs at the central position. However, slowest responses were observed at the right terminal position of the string, comparable to the results of the non-letters.

The linear effect for letters explains 88.2 % of variance. For shapes, there was a linear as well as a quadratic effect, which explains 91.7 % of variance. RTs and model fits are illustrated in Figure 2.

Discussion

In this study we investigated distinctive processing of letters and non-letters in five-character strings. In contrast to previous studies, by using strings which were randomly composed of Latin letters, Hebrew letters and Cyrillic letters (non-letters for the participants), we determined whether different processing strategies can be triggered by the target character, which had to be detected. As a result of the APPLE (Lachmann, in press), we expected that the analytic processing strategy (i.e. from left to right with a benefit for the terminal positions) would be automatically elicited, if the participants were instructed to detect a Latin letter target. Since the participants were not familiar with reading Hebrew or Cyrillic, a more holistic reading strategy (i.e. symmetrical scanning out from the central to the terminal positions) is expected if participants were instructed to detect a target out of these alphabets.

The results support our hypotheses. For RTs, a linear model explains 88.2 % of variance for Latin letters; for Hebrew and Cyrillic letters a symmetrical U-shaped position function, explaining 91.7 % of variance, was extracted. Since for all conditions the strings were composed of Latin letters as well as Hebrew letters and Cyrillic letters this distinction could not be caused by the string itself. This distinction could only result from the nature of the target character, suggesting that the pre-designated target triggers the searching strategy. Hence, if participants were instructed to detect a Latin letter in the string, a typical reading strategy was activated, i.e. character recognition and encoding from left to right. This idea, that reading or letter-specific processing strategies are elicited by the type of material which should be processed is in accordance with prior studies (Lachmann & van Leeuwen, 2004, 2008; van Leeuwen & Lachmann, 2004). For example, different effects where shown by using contextual embedded figures (e.g., surrounded letters and non-letter shapes), depending whether participants should make a categorical decision (letter versus non-letter; van Leeuwen & Lachmann, 2004) or if participants should decide in a same-different task whether two characters, either letters or non-letters, are the same (Lachmann & van Leeuwen, 2004, 2008). These findings support our hypothesis that prior knowledge about the target influences the participants' strategy. Prior studies demonstrated that differences of familiarity or meaning are not sufficient to explain different letter and non-letter processing (Lachmann et al., 2014; Lachmann & van Leeuwen, 2004; Schmitt et al., 2017; Lachmann, in press). Their results support the hypotheses of different processing strategies: an analytic processing strategy for letters versus a holistic object recognition strategy for non-letters. The results of the current study showed that the automatized analytic strategy elicited by letters is not limited to single-letter processing but further influences the processing of whole character-strings.

References

Burgund, E. D., Schlaggar, B. L. & Petersen, S. E. (2006). Development of letter-specific processing: The effect of reading ability. *Acta Psychologica, 122*(2006) 99–108. doi:10.1016/j.actpsy.2005.11.005

Coltheart, M., Curtis, B., Atkins, P. & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100*(4), 589–608. doi: 10.1037/0033-295X.100.4.589

Farid, M. & Grainger, J. (1996). How initial fixation position influences visual word recognition: A comparison of French and Arabic. *Brain and Language, 53*(3), 351–368. doi: 10.1006/brln.1996.0053

Fernandes, T. & Kolinsky, R. (2013). From hand to eye: The role of literacy, familiarity, graspability, and vision-for-action on enantiomorphy. *Acta Psychologica, 142*(1), 51–61. doi: 10.1016/j.actpsy.2012.11.008

Grainger, J., Tydgat, I. & Issel´e, J. (2010). Crowding affects letters and symbols differently. *Journal of Experimental Psychology: Human Perception and Performance, 36*(3), 673–88. doi: 10.1037/a0016888

Green, D. W., Hammond, E. J. & Supramaniam, S. (1983). Letters and shapes: Developmental changes in search strategies. *British Journal of Psychology, 74*(1), 11–16. doi: 10.1111/j.2044-8295.1983.tb01839.x

Hammond, E. J. & Green, D. W. (1982). Detecting targets in letter and non-letter arrays. *Canadian Journal of Psychology, 36*(1), 67–82. doi: 10.1037/h0081211

Ktori, M. & Pitchford, N. J. (2008). Effect of orthographic transparency on letter position encoding: A comparison of Greek and English monoscriptal and biscriptal readers. *Language and Cognitive Processes, 23*(2), 258–281. doi: 10.1080/01690960701536797

Ktori, M. & Pitchford, N. J. (2009). Development of letter position processing: Effects of age and orthographic transparency. *Journal of Research in Reading, 32*(2), 180–198. doi: 10.1111/j.1467-9817.2009.01394.x

Lachmann, T. (in press). Reading and dyslexia. The Functional Coordination Framework. In T. Lachmann, & T. Weis (eds.) *Reading and dyslexia*. Literacy Studies, 16. Cham: Springer.

Lachmann, T., Schmitt, A., Braet, W. & van Leeuwen, C. (2014). Letters in the forest: Global precedence effect disappears for letters but not for non-letters under reading-like conditions. *Frontiers in Psychology, 5*, 705. doi: 10.3389/fpsyg.2014.00705

Lachmann, T. & van Leeuwen, C. (2004). Negative congruence effects in letter and pseudo-letter recognition: The role of similarity and response conflict. *Cognitive Processing, 5*(4), 239–248. doi: 10.1007/s10339-004-0032-0

Lachmann, T. & van Leeuwen, C. (2008a). Differentiation of holistic processing in the time course of letter recognition. *Acta Psychologica, 129*(1), 121–129. doi: 10.1016/j.actpsy.2008.05.003

Lachmann, T. & van Leeuwen, C. (2007). Paradoxical enhancement of letter recognition in developmental dyslexia. *Developmental Neuropsychology, 31*(1), 61–77. doi: 10.1080/87565640709336887

Mason, M. (1982). Recognition time for letters and nonletters: Effects of serial position, array size, and processing order. *Journal of Experimental Psychology: Human Perception and Performance, 8*(5), 724–738. doi: 10.1037/0096- 1523.8.5.724

Pegado, F., Nakamura, K., Braga, L. W., Ventura, P., Nunes Filho, G., Pallier, C., . . . Dehaene, S. (2014). Literacy breaks mirror invariance for visual stimuli: A behavioral study with adult illiterates. *Journal of Experimental Psychology: General, 143*(2), 887–94. doi: 10.1037/a0033198

Pegado, F., Nakamura, K., Cohen, L. & Dehaene, S. (2011). Breaking the symmetry: Mirror discrimination for single letters but not for pictures in the visual word form area. *NeuroImage, 55*(2), 742–749. doi: 10.1016/j.neuroimage.2010.11.043

Pitchford, N. J., Ledgeway, T. & Masterson, J. (2008). Effect of orthographic processes on letter position encoding. *Journal of Research in Reading, 31*(1), 97–116. doi: 10.1111/j.1467-9817.2007.00363.x

Poirel, N., Pineau, A. & Mellet, E. (2008). What does the nature of the stimuli tell us about the Global Precedence Effect? *Acta Psychologica, 127*(1), 1–11. doi: 10.1016/j.actpsy.2006.12.001

Schmitt, A. (2016). *What makes a letter a letter? - New evidence for letter-specific processing strategies*. TU Kaiserslautern, Germany

Schmitt, A., van Leeuwen, C., & Lachmann, T. (2017). Connections are not enough for membership: letter/non-letter distinction persists through phonological association learning. *Acta Psychologica, 176,* 85-91

Stevens, M. & Grainger, J. (2003). Letter visibility and the viewing position effect in visual word recognition. *Perception & Psychophysics, 65*(1), 133–51. doi: 10.3758/BF03194790

van Leeuwen, C. & Lachmann, T. (2004). Negative and positive congruence effects in letters and shapes. *Perception & Psychophysics, 66*(6), 908–925. doi: 10.3758/BF03194984

Wagstaffe, J., Pitchford, N. & Ledgeway, T. (2005). Does central fixation account for medial letter facilitation in visual search? *Perception, 34*(Suppl.), 150. doi: 10.1177/03010066050340S101

POSTER SESSION III

THE OTHER´S PAIN: HOW MOTHERS PERCEIVE THEIR INFANT´S PAIN PERCEPTION

Talyta S. Araujo, Lucas A. Vargas, José A. L. Jesus, and Rosana M. Tristão
Faculty of Medicine, University of Brasilia, Brasilia, Brazil

Abstract

Parents' responses to the observation of their baby's reaction to a nociceptive stimulus have been described as having a substantial impact on the child's pain experience (Vervoot et al, 2011). The parental response in turn can be modulated by their level of catastrophizing thoughts. Catastrophizing is a concept related to the exaggerated negative orientation toward noxious stimuli, which has a biological basis, but also has been associated to the development of negative emotional reactions as stress, depression and anxiety (Sullivan and Bishop, 1995). This study investigated the relation between perceived maternal catastrophizing to her newborn painful experience, looking at factors as the mother´s own pain experience and her stress level. It was a cross-sectional exploratory study of a convenience cohort where participated 103 mothers aged 26.3 (±6.3) up to 2-3 days postpartum and their infants with 37.3 (±2.8) gestational age in weeks and elective for blood test by heel prick. After consent, all mothers were questioned if they were experiencing pain in breastfeeding and for how long and the level of this pain was measured by the Wong-Baker FACES® Pain Rating Scale. They also filled the Pain Catastrophizing Scale - Parental version (PCS-P), a sociodemographic protocol and the Holmes-Rahe Life Stress Inventory. Clinical variables of mother-infant dyads were collected from the medical reports. This study was conducted at the Maternal and Child Health Unit of the University of Brasilia, Brazil. The results showed that PCS-P score divided in low and high catastrophizers had impact of the number of days the mother had feeling pain on breastfeeding ($p=.021$). The total score of PCS-P was negatively correlated with maternal age ($r=-.30$, $p=.007$) and infants weight classification. One-way analyses showed that gestational stress had effect over PCS-P [$F(1,102)=6.892$, $p=.010$). This association is potentialized when the infant is small for gestational age and of female gender ($p<.001$). Logistic Regression pointed to significant effect of lower level of maternal education ($p=.05$), not be willing of the pregnancy ($p=.015$) and being a younger mother ($p=.008$) as explanation for higher levels of maternal parental catastrophization. In this study challenges to the maternal perception to the infant´s pain was linked to maternal higher catastrophizing thoughts and to the pain they felt during the breastfeeding, what was more pronounced in younger mothers, with lower education levels.

Pain can be characterized as a complex, plural phenomenon involving biological, affective, sensory and cognitive aspects (Dzau and Pizzo, 2014). As a result, there is a consensus that pain can only be understood from a biopsychosocial perspective which makes its study in non-verbal patients, such as newborn babies, a challenge (Damme et al, 2010). Catastrophizing is a concept related to the exaggerated negative orientation toward noxious stimuli, but also has been associated to the development of negative emotional reactions as depression and anxiety and in turn parental responses can have a substantial impact on the child's pain experience (Caes et al, 2011, Vervoot et al, 2011). Pain catastrophizing is characterized by a tendency to increase the value of the stimulus or threat of pain and to feel powerless in the context of pain. This is due to relative inability to inhibit pain-related thoughts in anticipation, during or after a painful event. Pain acts both in attracting our own attention and in that of others, whose responses may, in turn, influence those who have suffered the pain. This may be particularly important in the context of pediatric pain; children and adolescents are highly dependent on parental care (Golianu et al 2007, Severeijns et al, 2001). Parental reactions are not always adaptive. Thus, it is important to understand which parents engage in specific behaviors such as reaction to their child's pain, why, and what the consequences might be (Tristão et al, 2015). The objectives of this study were to test and evaluate the pain catastrophic scale in mothers of neonate infants and to verify its relation with maternal prenatal stress and mother´s own pain experience in breastfeeding.

Method

It was a randomized, observational, prospective, controlled study by the same individual, type before, during and after. This study was conducted at the Maternal and Child Health Unit of the University of Brasilia Hospital between March 2017 and January 2018. The subjects of the study were 103 women and their newborns, mothers aged 26.3 (±6.3) and their infants with 37.3 (±2.8) gestational age in weeks and elective for blood test by heel prick. Mothers should have been primiparous in the mid-term puerperium (up to 2-3 days postpartum), who conceived healthy, term infants through natural delivery. Women who present external factors that may interfere with the behavior of stress were excluded from participation. Also excluded were women younger than 18 and older than 30; gestation fewer than 37 weeks or greater than 41 weeks and 6 days; those who have had a high-risk pregnancy, obstetric condition or current medical condition. High-risk pregnancy or obstetric conditions were defined as placental abruption, severe preeclampsia, cesarean delivery, gestational diabetes, incompetent cervix, uterine anomaly, more than second-trimester abortion, multiple gestations, gestation conceived through assisted reproduction technology and large leiomyomas (>4cm). Newborns will also be subjects of the research. The newborn may have had only two painful events before heel puncture, including vitamin K injection and Hepatitis B vaccine.

Procedures

Patients without any of these adverse conditions at the time of admission were considered at low risk for adverse events and were offered recruitment for the study. The study was explained to the patients and those who agree signed a free informed consent form authorizing the collection of medical records, as well as collecting some personal data for data analysis. After consent, all participants were submitted to two protocols PCS-C (Caes et al, 2011) and Wong–Baker faces pain scale (Garra et al, 2010), to an inventory of gestational stress by Holmes-Rahe Stress scale (Holmes & Rahe, 1967) and to an inventary for sociodemographic data in order to ascertain socio-economic and psychological issues that are important in establishing a correlation with data obtained from PCS-P responses. This study was run in accordance with the regulations of the Ethics Committee of the Faculty of Medicine of the University of Brasília and it´s under the registration number: CEUnB-67983517.8.0000.5558.

The inventories used were: 1) Pain catastrophizing scale - parental version – PCS-P*:* This instrument is an adaptation of the Adult Pain Catastrophic Scale and the Catastrophic Pain Scale for Childhood. The PCS-P consists of 13 items that describe different thoughts and feelings that parents can present when their child is in pain. Parents rate how often they experience each of their thoughts and feelings when the child is in pain using a 5-point scale (0 = "nothing" to 4 = "extremely"). PCS-P consists of three subscales: (1) rumination (example: "... I keep thinking about how much it hurts"), (2) magnification (example, "... I wonder if something serious could happen"), and (3) helplessness facing the other´s pain (for example," ... there is nothing I can do to reduce pain") and produces a total score that can range from 0 to 52.*;* 2) Wong–Baker faces pain scale: The WBFS combines pictures of expressive faces and numbers to enable the user to rate pain. It can be used for children over the age of three and for adults. The faces range from a smiling face to a sad, crying face. A numerical rating is assigned to each face (from 0, "no hurt" to 10, "hurts worst"); 3) Maternal Prenatal Stress: For each participant, the Holmes & Rahe scale (1967) was administered. Regarding its impact on life, it is attributed a specific value to each stressor and the final score is classified into three categories: no stress (≤ 150 points); moderate level (151 to 299 points); high level (≥ 300 points).

We tested the normality of the variables using the Kolmogorov-Smirnov test. Because of the rejection of normality assumption of the clinical variables, we report those data as median and interquartile range and we performed non parametric tests. We used the Wilcoxon signed rank test and the Spearman correlation coefficient to compare the scores obtained with the scales and sociodemographic data. We also compared the scores of mothers with high and low level of catastrophization using the Wilcoxon sum rank test. To determine the relevance of the studied variables in relation to maternal catastrophization to pain, we performed a multivariate linear regression model. Statistical analyses were performed by using SPSS V.23.

Results

Following cut-off point of score 42 criterion that divided the sample between parental catastrophizers (>42) and no catastrophizers (≤42), our sample had 36 (35.5%) mothers with low level of catastrophization. The general description of the sample characteristics can be found at Table 1. For breastfeeding pain, 43 mothers (42%) referred feeling pain rating pain level 6 in average (±1.8) in the Wong-Baker Face scale and 14 (14%) complained of pain after the third day after birth; 53% breastfeed at the first hour of life of the baby and 12.6% after six hour after delivery; 71% considered breastfeeding a pleasant activity.

Pearson correlation analyses showed that total score of PCS-P was negatively correlated with maternal age (r=-.30, p=.007) and infants weight classification. One-Way analyses showed that Gestational stress had effect over PCS=P [F(1,102)= 6.892, p=.010) with significant correlation to the item related to persistent thinking of pain (F(1,102)= 3.142, p=.049). The results showed that PCS-P score divided in low and high catastrophizers had impact of the number of days the mother had feeling pain on breastfeeding (p=.021). The total score of PCS-P was negatively correlated with maternal age (r=-.30, p=.007) and infants weight classification. One-way analyses showed that gestational stress had effect over PCS-P increasing the exacerbation of the perception of other´s pain [F(1,102)=6.892, p=.010). This association is potentialized when the infant is small for gestational age and of female gender (p<.001). Logistic Regression pointed to significant effect of lower level of maternal education (p=. 05), not be willing of the pregnancy (p=. 015) and being a younger mother (p=.008) as explanation for higher levels of maternal-parental catastrophization of the pain perception in infants (Table 1).

Discussion and Conclusions

The beginning and the success of breastfeeding after delivery is usually described as an emotional joyful experience and it is linked to major advantages for infants' neurodevelopment. Conversely, mothers' or infants' clinical factors or other unexpected conditions can delay or disrupt or even prevent a health and joyful nursing and maybe linked to risk factors for maternal mental health and difficulties in caring the newborn baby. The parental perception of theirs infants´ pain response has been identified as one of the most important psychological variables to explain the magnitude of pain responses in clinical and non-clinical situations (Dzau and Pizzo, 2014, Damme et al, 2010).

In this study challenges to the maternal perception to the infant´s pain was linked to maternal higher catastrophizing thoughts and to the pain they felt during breastfeeding, what was more pronounced in younger mothers, with lower education levels. Also, birth weight and infants' gender posed themselves as risk factors when baby girls and lower weight were more linked to catastrophizing thoughts. Considerable research has shown that catastrophizing contributes to more severe pain, disability and increased emotional stress in response to pain. In addition, catastrophizing has been associated with increased use of health services, longer hospitalizations and use of medications (Caes et al, 2011). The neurophysiological mechanisms by which pain catastrophizing is related to acute and chronic pain are attracting research interest. Understanding these mechanisms has the unique potential to shed light on the main central nervous system factors that mediate pain catastrophic relationships and therapeutic benefits associated with changes in perceptual-cognitive related processes. The development and validation of tools capable of evaluating pain catastrophizing play an important role in the understanding of pain perception in infants.

Acknowledgements

We are very grateful to all parents who took part in this study. This research was supported by a grant from FAPDF - Research Support Foundation of the Federal District -FAPDF/ National Research Council - CNPq, linked to the Ministry of Science, Technology and Innovation of Brazil (grant number 193.000.324/2007) awarded to Rosana M Tristão.

Table 1. Sample characteristics for mothers-infants' sociodemographic and clinical data. Values are expressed as mean ± standard deviation ($\overline{M}$± SD) or frequency (%)

Sample Characteristics	(N=103)
Abdominal delivery [n (%)]	41 (39.8%)
Delivery Complications (yes)	27 (26.2%)
Weight at birth (M ± SD)	3001.7±740.8
Apgar 1st min (M±SD)	8.3± 1.0
Apgar 5th min (M±SD)	8.9± .7
Baby Gender - Female [n (%)]	43 (41.7%)
Maternal education [n (%)]	
Basic education	11 (9.7%)
Secondary	12 (11.7%)
High School	48 (46.6%)
Graduation	32 (31,1%)
Working during pregnancy - Yes [n (%)]	47 (45.6%)
Willing to be pregnant [n (%)]	
Unwilling	12 (11.7%)
Low willing	6 (5.8%)
Willing	16 (15.5%)
High willing	69 (67.0%)
Marital Status [n (%)]	
Single	18 (17.5%)
Stable union	41 (39.8%)
Dissolved union	3 (2.9%)
Family income	
1-3 minimum wages	79 (76.7%)
3-6 minimum wages	17 (16.5%)
Over 6 minimum wages	70 (6.80%)
Holmes-Rahe Stress scale (M ± SD)	
Total score	254.8±118.8
Low stress (N/%)	20 (19.4%)
Moderate stress	49 (47.6%)
High stress	33 (32.0%)
Parental Catastrophizing Scale total score (M ± SD)	48.8 ± 8.0
Magnification factor	10.9 ± 2.5
Rumination factor	17.3 ± 2.6
Helplessness factor	20.3 ± 4.2

References

Caes, L., Vervoort, T., Eccleston, C., Vandenhende, M., & Goubert, L. (2011). Parental catastrophizing about child's pain and its relationship with activity restriction: The mediating role of parental distress. *Pain*, 152 (1), 212-222.

Damme, S. V., Legrain, V., Vogt, J., & Crombez, G. (2010). Keeping pain in mind: A motivational account of attention to pain. *Neuroscience & Biobehavioral Reviews*, 34 ((2), 204-213.

Dzau, V. J., & Pizzo, P. A. (2014). Relieving Pain in America. *Jama*, 312 (15), 1507.

Garra G., Singer A. J., Taira B. R., Chohan J., Cardoz H., Chisena E, & Thode H. C. Jr. (2010). Validation of the Wong–Baker FACES Pain Rating Scale in pediatric emergency department patients. *Academic Emergency Medicine,* 17 (1), 50–54.

Golianu, B., Krane, E., Seybold, J., Almgren, C., & Anand, K. (2007). Non-Pharmacological Techniques for Pain Management in Neonates. *Seminars in Perinatology, 31* (5), 318-322.

Holmes T. H., & Rahe R. H., (1967). The Social Readjustment Rating Scale. *J Psychosom Res*, 11 (2): 213–218.

Severeijns, R., Vlaeyen, J. W., Hout, M. A., & Weber, W. E. (2001). Pain Catastrophizing Predicts Pain Intensity, Disability, and Psychological Distress Independent of the Level of Physical Impairment. *The Clinical Journal of Pain,* 17 (2), 165-172.

Sullivan, M. J. L., Bishop S. R., & Pivik J. (1995). The Pain Catastrophizing Scale: development and validation. *Psychol Assess*, 7, 524–32.

Tristão, R. M., Neiva, E. R., Barnes, C. R., & Adamson-Macedo, E. (2015). Validation of the Scale of Perceived Self-Efficacy of Maternal Parenting in Brazilian Sample. J. Hum. Growth Dev. *Journal of Human Growth and Development*, 25 (3), 277-286.

Vervoort, T., Caes, L., Crombez, G., Koster, E., Damme, S. V., Dewitte, M., & Goubert, L. (2011). Parental catastrophizing about children's pain and selective attention to varying levels of facial expression of pain in children: A dot-probe study. *Pain*, 152(8), 1751-1757.

MULTISENSORY INTERACTION BETWEEN OLFACTION AND FACE GENDER JUDGEMENTS

Patricia Geiger[1], Alice Stephan[1], Tina Plank[1], Géza G. Ambrus[2], Gyula Kovács[2], and Mark W. Greenlee[1]
1 Institute of Experimental Psychology, University of Regensburg, Germany
2 Institute of Psychology, Friedrich Schiller University Jena, Germany
PatriciaGeiger96@web.de, alice.stephan@ur.de, tina.plank@ur.de, geza.ambrus@uni-jena.de, gyula.kovacs@uni-jena.de, mark.greenlee@ur.de

Abstract

In this study, 24 participants had to judge the gender of faces presented to them for 200 ms in each trial in a 2AFC-task, while two odors were presented to them, alternating with an odorless solution, via an olfactometer. The two odors "cherry" and "mint" had been characterized as "female" and "male", respectively, in a pilot experiment. The presented faces were morphed between male and female faces with varying portions (20 to 80 % male) of each gender. Estimated psychometric functions based on the percentages of "male" responses for each morph level under each odor condition revealed that the mean 50%-threshold, i.e. the point of subjective equality between the perception of a face as male or female, decreased to a morph level of 39.9 % male under the odor condition "masculine" (mint) and showed a relative increase to a morph level of 43.0 % male under the odor condition "feminine" (cherry). Repeated-measures ANOVAs on the proportion of "male" responses revealed a marginally significant main effect of odor condition (p = .09), that was significant when the subgroup of female participants was analyzed separately (p = .018). The results point to an at least moderate effect of odors on face gender judgements, that is especially pronounced in female observers.

Studies on multisensory perception have already shown multiple cases of interactions between olfaction and vision, for example the influence of color on odor identification (e.g. Zellner et al., 1991), or the effects of olfaction on binocular rivalry (Zhou et al., 2010) or grasping behavior (Tubaldi et al., 2008).

In this study we investigated the effect of congruent and incongruent odors on face gender judgements. In a pilot study we identified two odors (cherry and mint) out of ten presented odors, which 19 subjects significantly characterized as "feminine" and "masculine", respectively. Subsequently, we presented these two odors in a face gender judgement task.

Method

We presented the two odors "cherry" and "mint", alternating with a neutral odorless solution, in a randomized order to 24 participants (16 females, 8 males; age: 18 to 59 years) via an olfactometer, while they performed a face gender judgement task. In this 2AFC-task on each trial a morph (colored photograph) out of a female and a male face (morph levels 20, 30, 35, 40, 45, 50, 55, 60, 65, 70 and 80 % male) was presented for 200 ms (Fig. 1), and the participants had to indicate by button press, if they perceived the face as male or female. We calculated psychometric functions for each odor condition, indicating the percent of "male" responses in relation to each morph level, and additionally compared the mean proportions of "male" responses for each morph level and odor condition with a repeated-measures ANOVA.

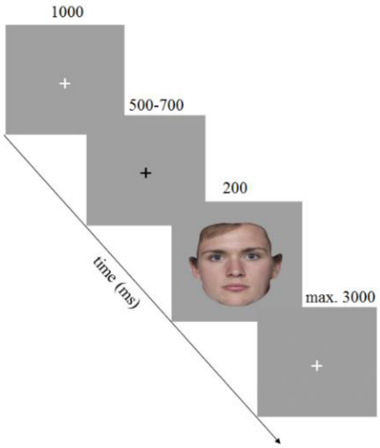


Fig. 1. Depiction of a typical trial in the 2AFC-face-gender-judgement task. After a variable interval of 500 to 700 ms, a face with varying morph level between "male" and "female" (here: % male) was presented for 200 ms. Afterwards, the participants had maximum 3 seconds time to give their response via button press. Each block, in which one odor was presented, consisted of eleven trials with different face stimuli. Odor blocks were alternated by blocks, where a neutral odorless solution was presented.

Results

The results show that the mean 50%-threshold, i.e. the point of subjective equality between the perception of a face as male or female, lies at a morph level of 39.9 % male under the odor condition "masculine" (mint) and increased moderately to a morph level of 43.0 % male under the odor condition "feminine" (cherry) (n=23 here, one subject excluded because fit of psychometric function not possible). The repeated-measures ANOVA on the proportions of "male" responses with the factors odor condition ("masculine", "feminine") and morph level (20-80% male) revealed a marginally significant main effect of odor condition [$F(1,23) = 3.1$; $p = .09$] and a significant main effect of morph level [$F(10,230) = 84.4$; $p < .001$; Greenhouse-Geisser corrected]. When we analyzed the data of male (n=8) and female (n=16) participants separately, we could observe that the effect of odor condition was mainly driven by the female participants. The female group showed a significant main effect of odor condition [$F(1,15) = 7.1$; $p = .018$], a significant main effect of morph level [$F(10,150) = 75.1$; $p < .001$] and a significant interaction between odor condition and morph level [$F(10,150) = 2.0$; $p = .037$]. On the other hand, the male group showed only a significant main effect of morph level [$F(10,70) = 17.6$; $p < .001$]. Fig. 2 shows the mean psychometric functions for the two odor conditions "masculine" (mint) and "feminine" (cherry), separately for the group of female participants and for the group of male participants.

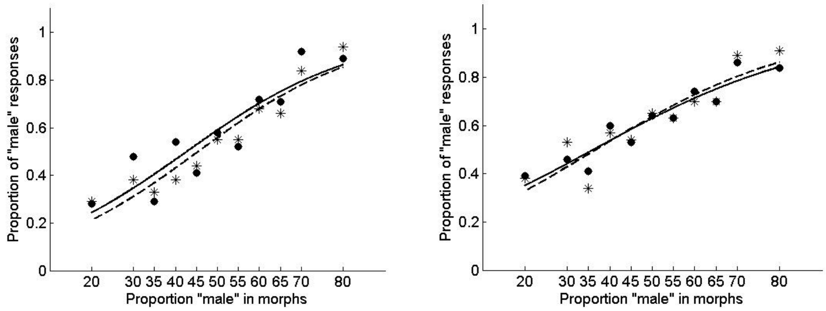


Fig. 2. Mean psychometric functions based on the portions of „male“ responses as a function of morph levels (20 to 80 % male). Solid lines and black circles show the psychometric functions and data obtained under the odor condition “mint” (= odor judged as “masculine”), dashed lines and asterisks show the psychometric functions and data obtained under the odor condition “cherry” (= odor judged as “feminine”). Left panel: mean values for the female participants only. Right panel: mean values for the male participants only.

Discussion

The results suggest an at least moderate effect of odors characterized as “male” or “female” on face gender judgements. Female observers appear to be especially prone to this effect. This points to an interesting difference in the processing of multisensory stimuli between men and women, and possibly to a stronger functional connection between visual and olfactory brain areas in women. Indeed, Robinson et al. (2016) could show that female observers had an advantage in a visual object recognition task when the masked objects were accompanied by congruent odors. The male observers in that study did not show that effect. Additionally, an EEG study by Robinson et al. (2015) found that the N1 component of the visual event-related potential was enhanced, when presented odors matched presented visual objects, but this was again only found in women, not in men. In a follow-up experiment, with an extended set of face stimuli, we plan to investigate gender differences in multisensory judgments of gender identity in more detail.

References

Robinson, A. K., Laning, J., Reinhard, J., & Mattingley, J. B. (2016). Odours reduce the magnitude of object substitution masking for matching visual targets in females. *Attention, Perception, & Psychophysics*, **78**(6), 1702-1711.

Robinson, A. K., Reinhard, J., & Mattingley, J. B. (2015). Olfaction modulates early neural responses to matching visual objects. *Journal of cognitive neuroscience*, **27**(4), 832-841.

Tubaldi, F., Ansuini, C., Tirindelli, R., & Castiello, U. (2008). The grasping side of odours. *PLoS One*, **3**(3), e1795.

Zellner, D. A., Bartoli, A. M., & Eckard, R. (1991). Influence of color on odor identification and liking ratings. *The American journal of psychology*, **104**(4), 547-561.

Zhou, W., Jiang, Y., He, S., & Chen, D. (2010). Olfaction modulates visual perception in binocular rivalry. *Current Biology*, **20**(15), 1356-1358.

EFFECTS OF SENSITIVITY, BIAS, AND STIMULUS PRESENTATION ORDER IN COMPARATIVE DISCRIMINATION OF INTERVAL DURATIONS: A UNIVARIATE AND MULTIVARIATE STUDY

Åke Hellström
Stockholm University, SE-106 91 Stockholm, Sweden
Thomas H. Rammsayer
University of Bern, Switzerland
hellst@psychology.su.se, thomas.rammsayer@psy.unibe.ch

Abstract

According to the sensation weighting model (SWM), stimulus magnitude level dependent time-order errors (TOEs) in stimulus comparison arise from uneven stimulus weighting, levered by asymmetry of an internal reference level; the weighting also causes discriminability to depend on the presentation order of standard (St) and comparison stimulus (Co) (the Type B effect). Both of these effects, as well as judgment bias, determine the measured difference limen (DL). In two duration discrimination experiments, these contributions to the DL were explored, using an adaptive staircase method. The compared intervals were filled auditory or empty visual. The interstimulus interval was 900 ms and the St duration 100, 215, 464, or 1000 ms in a blocked design. In univariate as well as multivariate analyses, the SWM's predictions were confirmed on the individual level, and the contributions of sensitivity, bias, and weighting in building the DL were assessed.

Hellström's (1979) sensation weighting model (SWM) can be written,

$$d_{12} = [s_1 \cdot \psi_1 + (1 - s_1)\,\psi_{r1}] - [s_2 \cdot \psi_2 + (1 - s_2)\,\psi_{r2}] + b, \qquad (1)$$

where d_{12} is the subjective difference between the 1st and the 2nd stimulus. ψ_1 and ψ_2 are the sensation magnitudes of the stimuli, s_1 and s_2 are weights, and ψ_{r1} and ψ_{r2} are the reference levels (ReLs). The ReL is formed by pooling of stimulus information (Helson, 1964). It is normally near the center of the stimulus range, and it may differ between the stimuli. b is judgment bias. With two-alternative forced choice (2AFC), participants must guess when uncertain, which may yield "indecision bias" (García-Pérez & Alcalá-Quintana, 2017). When two equal stimuli ($\psi_1 = \psi_2 = \psi$) are compared, we get (for $\psi_{r1} = \psi_{r2} = \psi_r$),

$$d_{12} = (s_2 - s_1)\,(\psi_r - \psi) + b, \qquad (2)$$

where a value of $d_{12} \neq 0$ is the same thing as a *time-order error* (TOE).

We used an adaptive staircase method, the weighted up-down method (Kaernbach, 1991) to measure the difference limen (DL) for the duration of successive intervals (Exp. 1: filled auditory–white noise; Exp. 2: Empty visual–red LED light). Standard (S) durations were 100, 215, 464, and 1000 ms, and the interstimulus interval (ISI) was 900 ms.

Method

Durations of two successive intervals, S and C, were compared using 2AFC response. There were four pairs of 64-trial blocks, each block pair with one Higher- and one Lower-block. In a Higher- (Lower-) block, the Higher (Lower) 75% difference threshold (DL) was estimated, and was free to attain nonpositive values. Accuracy was emphasized. No feedback was given. The intertrial interval was 900 ms. The raw DL (rDL) was the mean, across the last 20 trials, of C – S (in Higher blocks) or S – C (in Lower blocks).

It was assumed that the psychophysical function was linear, $\psi = \phi$, and that the ReL could differ between Higher- (H) and Lower- (L) blocks. With *BT* = block type,

$$d_{12} = s_1\phi_1 - s_2\phi_2 + (s_2 - s_1)\, ReL_{BT} + b, \qquad (3)$$

For the *i*th participant, at the threshold d_{12} was assumed to be the proportion w_i of the subjective duration of S. The *raw DL* (rDL) then is $w_i \cdot$ S. Judgment bias was modeled as $b_i \cdot$ S.

The *raw Weber fraction* (rWF) was computed as rDL/S. A pair was called Up (U) when the 1st stimulus was initially shorter than the 2nd, and Down (D) in the opposite case. Thus, pair types were SCU, SCD, CSU, and CSD. The equations become:

$$\textit{Higher-block}, \text{ SCU: } rWF_{SCU} = (1/s_2)\,[w + b + (s_2 - s_1)\, RHQ]; \qquad (4)$$

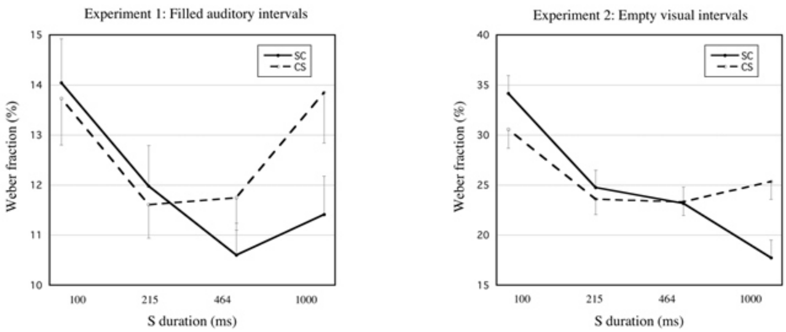
$$\textit{Lower-block}, \text{ SCD: } rWF_{SCD} = (1/s_2)\,[w - b - (s_2 - s_1)\, RLQ]; \qquad (5)$$

$$\textit{Lower-block}, \text{ CSU: } rWF_{CSU} = (1/s_1)\,[w + b + (s_2 - s_1)\, RLQ]; \qquad (6)$$

$$\textit{Higher-block}, \text{ CSD: } rWF_{CSD} = (1/s_1)\,[w - b - (s_2 - s_1)\, RHQ], \qquad (7)$$

where *RHQ* or *RLQ* is the relative displacement of the ReL from the S: $RHQ = (ReL_H - S)/S$ and $RLQ = (ReL_L - S)/S$. With RHQ = RLQ = RMQ, $WF_{SC}/WF_{CS} = s_1/s_2 = q$. With $q < 1$, the *Type B effect* (TBE) (Bausenhart et al., 2015) is negative, and with $q > 1$, it is positive.

ANOVA results for WFs. As is predicted by the SWM when the weighting $(s_2 - s_1)$ changes with the S duration, the ANOVA of WFs for Exp. 2 showed a significant interaction of S Duration X Order, $p = .003$, to which only the linear effect of S Duration contributed significantly, $p < .001$. For S = 1000 ms, $WF_{CS} > WF_{SC}$, $p = .002$ (a neg. TBE). For S = 100 ms, WF_{SC} SC > WF_{CS}, $p = .042$ (a pos. TBE). Bonferroni corrected *p*s were .006 and .17. For Exp. 1, the interaction of S Duration X Order was nonsignificant, $p = .076$. Still, the effect of the linear component of S Duration was significant, $F(1,60) = 7.445$, $p = .008$, $\eta^2_p = .110$.



	Exp. 1				Exp. 2			
Est. s_1/s_2	1.02	1.03	0.90	0.82^{***}	1.12	1.05	0.99	0.59^{***}

	Exp. 1	Exp. 2
S duration	$p = .004$, $\eta^2_p = .200$	$p < .001$, $\eta^2_p = .526$
Order (SC,CS)	*ns*	*ns*
RMQ lev. X S dur.	*ns* (lin. $p < .001$)	$p = .003$, $\eta^2_p = .260$ (lin. $p < .001$)

Fig. 1. Mean Weber fraction (in %) for orders SC and CS, versus (S) duration (log time scale) for Exps. 1 and 2. SE_M bars, estimated values of s_1/s_2, and ANOVA results are also shown.

Time-Order Error (TOE)

A positive TOE means that the 1st duration is overestimated relative to the 2nd, so that a greater difference between C and S is needed to detect this difference with an Up than with a Down profile. The TOE is defined as $(DL_U - DL_D)/2$ and the TOE quotient (QTOE), TOE/S, as $(WF_U - WF_D)/2$. The mean predicted QTOE becomes:

$$QTOE_M = ½\,(QTOE_{SC} + QTOE_{CS}) = ½\,[b\,(s_1+s_2)/s_1s_2 + RMQ\,(s_2^2-s_1^2)/s_1s_2]. \quad (8)$$

$$QTOE_{SC} - QTOE_{CS} = b\,(s_1-s_2)\,/\,s_1s_2 - RMQ\,(s_1-s_2)^2\,/\,s_1s_2, \quad (9)$$

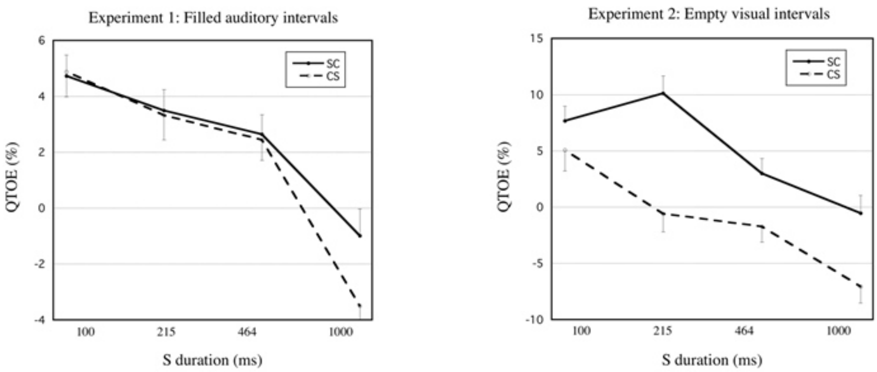
$QTOE_{SC}$ and $QTOE_{CS}$ are shown in Fig. 2. Their generally positive difference suggests $RMQ < 0$ (Hellström, 1979, 2000; Hellström & Rammsayer, 2015; Helson, 1964), as does the relation between QTOE and s_1/s_2, $r = .90$ (Exp. 1, $r = .94$, Exp. 2). For $s_1/s_2 \approx 1$, $QTOE_M > 0$, indicating a positive bias (i.e., toward "1st interval longer.")

Multivariate Approach: Principal Components Analyses

The interindividual variability of the rWFs carries information that is lost in univariate statistics. Therefore, we used a multivariate method to analyze the rWFs as well as the QTOEs. The two sets of results were consistent. Full details are given in Hellström and Rammsayer (submitted). Here we discuss the analyses of the QTOEs.

Bias (b_i), and leverage (RMQ_i) were assumed participant (i-)-specific, and bias expression, β_k, and weighting ($s_{2k}-s_{1k}$) condition (k-) specific. The predicted QTOE for (i, k) is:

$$QTOE_{ik} = \beta_k\, b_i + (s_{2k}-s_{1k})\, RMQ_i. \quad (10)$$



	Exp. 1				Exp. 2			
SC≠CS	*ns*	*ns*	*ns*	*ns*	*ns*	$p<.001$	$p=.004$	$p=.009$

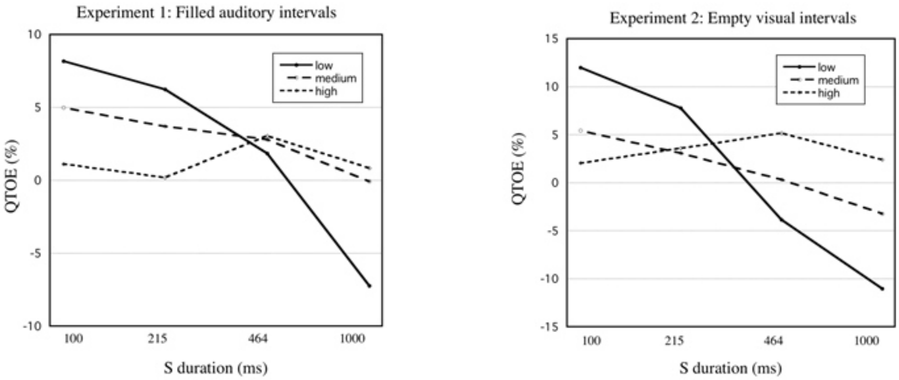
	Exp. 1	Exp. 2
S duration	$p < .001$, $\eta^2_p = .430$	$p < .001$, $\eta^2_p = .362$
Order (SC, CS)	*ns*	$p < .001$, $\eta^2_p = .507$
RMQ level X S dur.	*ns*	$p = .002$, $\eta^2_p = .266$

Fig. 2. Mean TOE quotient (QTOE, in percent) for orders SC and CS, plotted against standard (S) duration (log time scale) for Exps. 1 and 2; SE_M bars, ANOVAs, and *t* tests.

Values of $QTOE_{SC}$ and $QTOE_{CS}$ were entered into a principal components analysis. For each experiment, two components were extracted. The 1st component was identified with *b*, and the 2nd with *RMQ*. The mean values of β_k, as estimated from the loadings of the 1st component, confirmed the generally positive mean bias. Assuming $RMQ < 0$, the loadings of the 2nd component indicate $s_1/s_2 > 1$ for 100 and 215 ms, and $s_1/s_2 < 1$ for 464 and 1000 ms. The blocked design prevented large ReL asymmetries, limiting the leverage. Still, with *w* variability eliminated, the SW mechanism accounted for 20 - 30% of the variance.

Participants' RMQ component scores were partitioned at their low, medium, and high tertiles. RMQ scores modulated the slope of QTOE against S duration (see Fig. 3). This confirms the SWM's prediction that QTOE is proportional to RMQ multiplied by $(s_2 - s_1)$.

The positive correlations between RMQ score and QTOE for S = 1000 ms imply $(s_2 - s_1) > 0$; the negative correlations for, in particular, S = 100 ms in both experiments, imply $(s_2 - s_1) < 0$. So, it was confirmed that the weighting balance reversed into $s_1/s_2 > 1$ for brief S durations; significantly so for S = 100 ms (Exps. 1 and 2) and also for S = 215 ms (Exp. 1).



	Experiment 1	Experiment 2
r	$-.575^{***}$ $-.426^{***}$ $.205^{ns}$ $.575^{***}$	$-.447^{**}$ $-.128^{ns}$ $.599^{***}$ $.796^{***}$
RMQ level	$p = .534$, $\eta^2_p = .021$	$p = .524$, $\eta^2_p = .027$
S duration	$p < .001$, $\eta^2_p = .640$	$p < .001$, $\eta^2_p = .574$
RMQ lev. X S dur.	$p < .001$, $\eta^2_p = .349$	$p < .001$, $\eta^2_p = .302$

Fig. 3. Mean TOE quotient (QTOE, in percent) for each standard (S) duration at low-, medium-, and high-third score levels of *RMQ* component, as extracted from QTOEs. Included are ANOVA results. Correlation between *RMQ* component and QTOE is also given for each S duration (Bonferroni corrected: $^{***}p < .001$, $^{**}p < .01$, ^{ns} not significant).

Discussion

Our multivariate analyses of the TOEs identified bias and stimulus weighting as factors in their build-up. The decrease of s_1/s_2 with increasing stimulus duration parallels its decrease with increasing ISI that generally occurs in TOE experiments (e.g., Hellström, 1979, 2003). This change is thought to reflect the tuning of a mechanism that increases discrimination sensitivity by weighting-in of ReL information (Hellström, 1985; Patching et al., 2012) – a change of dominance, with longer ISIs, from stimulus interference to memory loss.

Ellinghaus et al. (2018) state that reported TBEs are negative except in rare cases with brief stimuli and short ISIs. However, the TBE is a rather insensitive indicator of the sensation-weighting balance, which is heavily dependent on the stimulus conditions. The multivariate analyses provided the statistical power needed to confirm the reversal of the weighting pattern, yielding $s_1/s_2 > 1$ for brief S durations (cf. Hellström, 1979, 1985, 2003; Patching et al.). Bias contributes to the interindividual TOE variation, but it does not account for the existence of the TOE or its variation across S durations and presentation orders.

References

Bausenhart, K. M., Dyjas, O., & Ulrich, R. (2015). Effects of stimulus order on discrimination sensitivity for short and long durations. *Attention, Perception & Psychophysics, 77*, 1033-1043.

García-Pérez, M. A., & Alcalá-Quintana, R. (2017). The indecision model of psychophysical performance in dual-presentation tasks: Parameter estimation and comparative analysis of response formats. *Frontiers in Psychology*, 12 July 2017.

Hellström, Å. (1979). Time errors and differential sensation weighting. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 460-477.

Hellström, Å. (1985). The time-order error and its relatives: Mirrors of cognitive processes in comparing. *Psychological Bulletin, 97*, 35-61.

Hellström, Å. (2000). Sensation weighting in comparison and discrimination of heaviness. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 6-17.

Hellström, Å. (2003). Comparison is not just subtraction: Effects of time- and space-order on subjective stimulus difference. *Perception & Psychophysics, 65*, 1161-1177.

Hellström, Å., & Rammsayer, T. H. (2015). Time-order errors and standard-position effects in duration discrimination: An experimental study and an analysis by the sensation-weighting model. *Attention, Perception & Psychophysics, 77*, 2409-2423.

Hellström, Å., & Rammsayer, T. H. (submitted). *Separate and joint effects of sensitivity, judgment bias, and presentation order in comparative discrimination of interval durations*.

Helson, H. (1964). *Adaptation-level theory*. New York: Harper & Row.

Kaernbach, C. (1991). Simple adaptive testing with the weighted up–down method. *Perception & Psychophysics, 49*, 227–229.

Michels, W. C., & Helson, H. (1954). A quantitative theory of time-order errors. *American Journal of Psychology, 67*, 327-334.

Patching, G. R., Englund, M. P., & Hellström, Å. (2012). Time- and space-order effects in timed discrimination of brightness and size of paired visual stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 915–940.

CROSMODAL ASSOCIATIONS BETWEEN OLFACTION AND VISION: COLOR AND SHAPE VISIUALIZATIONS OF ODORS

Kathrin Kaeppler
Institute of Experimental Business Psychology, Leuphana University Lueneburg, Wilschenbrucher Weg 84a, 21335 Lueneburg, Germany
kaeppler@leuphana.de

Introduction: In the present study, we assessed crossmodal associations between odors and both color and shape, with particular interest in the principles beneath these mappings. We hypothesized visual associations of odors would primarily reflect observable features of a smelling object and thus vary with different source assumptions of the very same smell.

Methods: We asked 30 participants to visualize their odor associations on a drawing tablet, freely deciding on color and shape. Additionally, subjects provided ratings on perceptual and shape–related dimensions as well as a verbal label for each sample.

Results: With respect to color selection, the results confirmed a source–based mapping approach: Odors rated as familiar were associated with very particular colors that typically resembled the appearance of their source. For less familiar odors, color selection was rather inconsistent but still then went along with assumed odor objects. Shape ratings changed with odor identifications as well, but considerably less than for color associations. Shape ratings and shape drawings produced very different results. While shape ratings were unlikely rooted in the mental imagery of a shape, drawings frequently displayed concrete objects that depended on odor label.

Conclusions: Results confirm the existence of stable odor-vision-correspondences and suggest that language plays a major part in mediating these mappings. The frequently assumed hedonic foundation of crossmodal matchings could not be confirmed for this stimuli set.

Implications: Odor sensations may trigger odor naming spontaneously. Assumptions about an odor's identity as well as the multisensory knowledge we have acquired on it, affect the visual associations of an odor.

QUANTIFYING ENVIRONMENTAL INTOLERANCE WITH A SMARTPHONE APP

Gregory Neely, Oskar Nilsson, Simon Lind, and Steven Nordin
Department of Psychology, Umeå University

Abstract

Environmental intolerance (EI) is a condition characterized by low tolerance to environmental stimuli at levels that would not affect most people. EI is an ill-defined condition from which sufferers experience highly individual multisystem symptoms following exposure from specific environmental sources. Most research on EI is conducted using cross sectional approaches, however, longitudinal approaches that capture daily exposure are needed to fully understand how EI develops and change over time. This paper describes an app that was developed that can be used with most smartphones and, in conjunction with a website, can be used to collect symptoms and ratings of discomfort in the field as well as qualitative reports of the incident that triggered the discomfort.

Sensitivity to the physical environment can have serious and debilitating consequences. Negative physiological and psychological effects usually accompany reactions to extreme sensory stimulation, such as high intensities or prolonged exposure. For example, even a very brief exposure to an extremely loud noise may cause permanent hearing loss. In both the work place and in everyday life, injuries and illnesses related to exposure to physically harming environmental agents such as chemicals, noise, and vibration have been well documented and researched (see, for example, Gan, Davies, & Demers, 2011; Griffin, Bovenzi, & Nelson, 2003; Dalton & Jaén, 2010).

However, there are individuals who display sensitivities to various environmental agents even at exposure levels that do not usually affect most people. This disorder is commonly referred to as environmental intolerance (EI) and is characterized by varied and multisystem symptoms that the individual attributes to specific environmental exposures (Sparks, 2000). This includes exposure to odorous chemicals, certain buildings, everyday sounds and electromagnetic fields (EMF). A recent survey with 3406 respondents from a county in northern Sweden found that slightly more than one fifth of the respondents reported being intolerant to odorous/pungent chemicals, certain buildings, EMF and/or sounds and that 6.3% of the responders had even had at least one intolerance diagnosis by a physician (Palmquist, Claeson, Neely, Stenberg & Nordin, 2014).

Common for all types of EI is that it is often difficult to establish physiological basis for reactions and that there is a large overlap between symptoms observed in the different intolerances. This has led some to theorize that there may in fact be a general environmental sensitivity in certain individuals. For example, some have suggested neural sensitization to sensory stimuli may be part of the problem (Anderson et al. 2009). Aron and Aron (1997) have proposed that proneness to overstimulation from environmental stimuli is a trait that is related to emotions and empathy. Binder and

Campbell (2004) suggest the medically unexplained symptoms characterizing are somatic stress disorders resulting from physical and/or psychological stressors. A study by Nordin, Ljungberg, Claeson, and Neely (2013) found that individuals that scored high on the Noise Sensitivity Scale (Weinstein, 1978) also scored high on a Perceived Stress Questionnaire (Levenstein et al., 1993) and on the Chemical Sensitivity Scale (Nordin, Millqvist, Löwhagen, & Bende, 2003).

Indeed, studies have found there to be overlaps between intolerances reported by individuals. Levallois, Neutra, Lee, and Hristaova (2002) found that of those reporting intolerance to EMFs 60.3% also reported intolerance to chemicals. Among individuals reporting intolerance to chemicals, 8.4% also reported intolerance to EMFs. Palmquist et al. (2014) found in a large population based study that 20.4% of the respondents that self-reported intolerance to chemicals also reported intolerance to sounds; and of those with a chemical intolerance diagnoses from a physician 18.1% had received a diagnoses of intolerance to sound. Conversely, 27.2% of the respondents that self-reported intolerance to sounds also reported intolerance to chemicals; and of those with a sound intolerance diagnoses from a physician 20.8% had received diagnoses of intolerance to chemicals. In fact, the same study found that the overlap of self-reported and diagnosed intolerance to one or more types of EI was significantly higher than could be expected by chance.

A large problem, however, is that EI largely has been examined using cross-sectional or experimental designs which makes it difficult to discern how these disorders develop over time and interact with each other and with the individual's environment and life situation. Using a postal survey with a 5-year follow-up Eek, Karlson, Österberg and Östergren (2010) found that participants who developed environmental annoyances between the two measurement periods reported more subjective health complaints, higher levels of stress and more job dissatisfaction at the initial measurement time compared to participants not developing environmental annoyance. However, this study used very simple measures for many of their variables (single sentence survey questions) and did not follow the participants continually through the time-period.

One solution to acquiring data over long periods and that reflect the experience of individuals as they go about their daily lives is collecting data with survey tools in smartphones. Naturally, this solution would not work well with those reporting intolerance to EMFs, but other types of EI should work.

We have developed an app (see Fig. 1) that will allow individuals to make quick reports on symptoms and ratings of discomfort on their smartphone at the time of the event with then the option of logging on to the record later to answer follow-up questions and provide more detailed information about the incident. Information can be put into the system even if the individual does not have direct internet access; the system saves the data until the phone is once again connected to the internet and then sends the data.

The app has a down-scaled, easy to use interface. Each report requires the user to go through three pages with seven questions. The total amount of time required for one report is approximately 45-60 seconds. The first page of the app includes three questions (Fig. 1, first panel); time of event, type of exposure and rating of discomfort associated to the exposure on a scale 1-10. The second page asks the user for symptoms associated with the exposure. The list of prespecified symptoms are based on the Environmental Hypersensitivity Symptom Inventory (EHSI) (Nordin, Palmquist, Claeson, & Stenberg, 2013). Symptoms are presented in six categories that, upon action, reveal further symptoms belonging to each category (see Fig. 1, second and third panels). On the third and final page (Fig. 1, fourth panel), users are asked to report where the event took place. They were also encouraged to provide keywords associated to the event in a free-text field. This option was included to ensure that details of events would be remembered later during follow up. To follow up an event, the respondent is required to log in to their account on a computer. Once logged in, the user can see the symptoms they reported and report how long the symptoms persisted (three-choice, category response scale; Less than 1 hour, 1-8 hours, More than 8 hours).

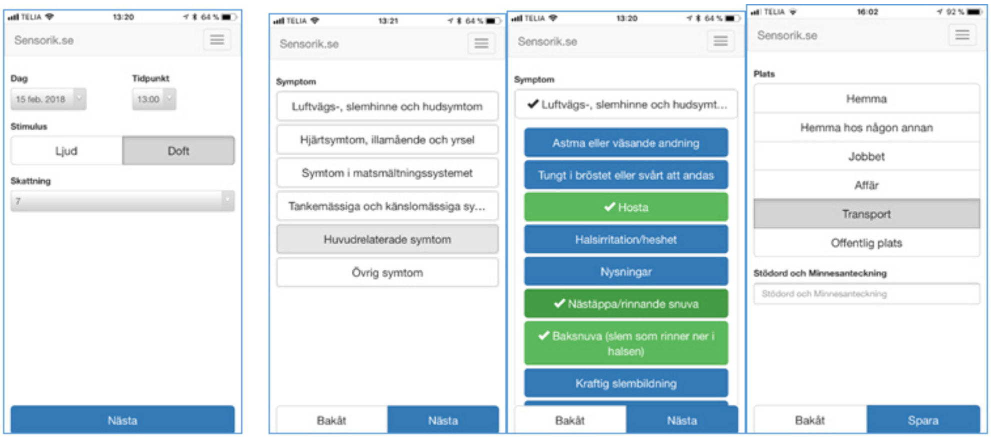


Fig. 1. The far left panel shows the initial screen. After filling in this screen the participant is taken to a screen where they can choose symptom groups (second panel from the left). Symptom panels can expand to show specific symptoms (third panel from the left). The final screen (right panel) allows the participant to indicate location and provide some key words in order to help recall the event later if they want to log onto the website to complement or adjust their responses.

Additionally, the user is provided with a free text field to complement their report with more detailed information about the event.

In order to test the functionality of the app, a small pilot study was conducted where individuals were asked to use the app during a two-week period to document events where they experienced discomfort from sound or smell exposures during the course of their daily living. After the data collection period, individuals were asked to complete environmental sensitivity questionnaires related to noise and chemical sensitivity. The sample is too small to statistically test difference in reporting from high and low sensitive individuals, but it is nevertheless anticipated that the number of reports of exposure, the number of symptoms reported and average ratings of discomfort will reflect at least a trend showing that higher sensitive individuals will report more exposures, more symptoms and higher ratings of discomfort.

Method

Twelve individuals (mean age 29 years, SD=10.7 years) were recruited from the local community. Participants received 100 SEK in compensation.

Participants were given written and oral instructions on how to install the app on their phones and instructed to report through the app every time they experienced an odor or noise they found unpleasant. The Noise Sensitivity Scale-11 (NSS-11; Nordin, Palmquist, & Claeson, 2013) and Chemical Sensitivity Scale for Sensory Hyperreactivity (CSS-SHR; Nordin et al., 2003) were completed by all participants after two weeks of using the app in order to have a standardized measure of their intolerance. Optimal cut-off scores for the NSS-11 and CSS-SHR have previously been developed to classify individuals as intolerant and non-intolerant (Nordin, Karvala, Nyback, & Sainio, 2018), but for the purpose of this pilot study, a relative cut-off score was calculated to divide participants into either high- or low intolerance (6 in each group for each sensitivity).

Results and Discussion

Participants were grouped into a high or low intolerance group based on their scores on the NSS-11 and CSS-SHR conditions, respectively (NSS low n=6, NSS high, n=6, CSS low n=6, CSS high n=6). Norm data for the general population have a mean score of 29.7 (SD=8.73) on the CSS-SHR with approximately normal distribution (Nordin, Palmquist, Bende, & Millqvist, 2013) . The same number for the NSS-11 is 27.2 (SD=7.97) (Nordin, Palmquist, & Claeson, 2013). A comparison between the sample in present study and norm data revealed that the low intolerance group corresponded to the 59th percentile of the population for the NSS-11 (mean score 29, SD=5.33) and the 30th percentile for the CSS-SHR (mean score 25.17, SD=4). The high intolerance group corresponded to the 88th and the 97th percentile of the population for the NSS-11 (mean score 42.67, SD=3.38) and CSS-SHR (mean score 40.17, SD=7.16) respectively.

A comparison between the total sample for the CSS-SHR and norm data revealed that present sample was within the 63rd percentile of the population (mean score 32.67, SD=9.60). Corresponding analysis for the NSS-11 showed that the sample was within the 86th percentile of the general population (mean score 35.83, SD=8.31).

As can be seen in Figure 2, as predicted the higher sensitive group, for both modalities, reported higher ratings of discomfort, more reports of incidents and more symptoms. As the sample is small, no statistical differences were observed but the trend was consistent over all variables and in both modalities. Few difficulties were reported from users regarding the actual functioning of the app, despite receiving relatively brief instruction (all respondents received written instructions, in some cases oral instructions were given in person and with other over the phone). The results seen in this small pilot indicate that the app has the potential to be a useful tool for gathering field measurements.

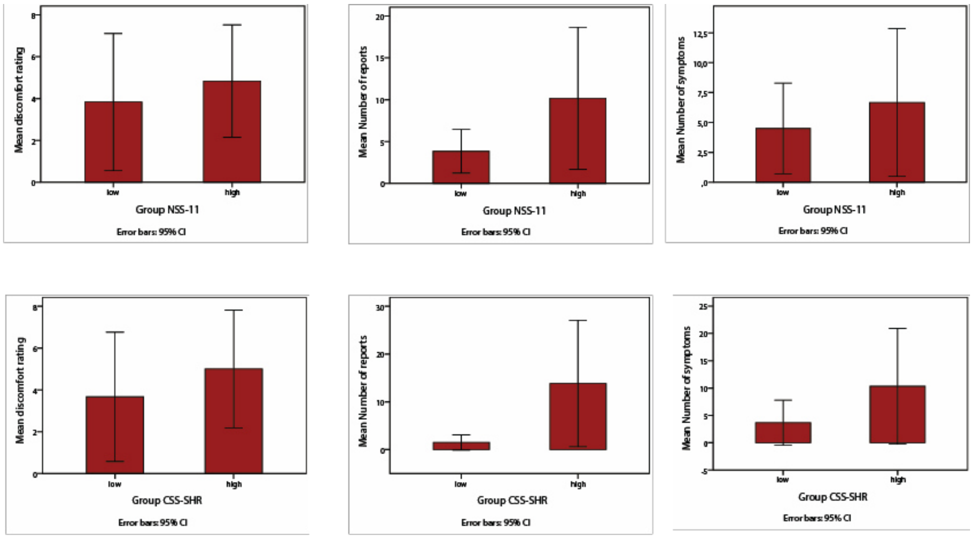


Fig. 2. Noise (top row) and odor (bottom row) responses to discomfort ratings, mean number of reports of incidents and mean number of reported symptoms for low sensitive (n=6) and high sensitive (n=6) respondents.

Acknowledgements

This research was sponsored by a grant to the first author from the Swedish Asthma and Allergy Association Research Foundation.

References

Andersson, L., Bende, M., Millqvist, E., & Nordin, S. (2009). Attention bias and sensitization in chemical sensitivity. Journal of Psychosomatic Research, 66(5), 407–416.

Aron, E. & Aron, A. (1997). Sensory-processing sensitivity and its relation to introversion and emotionality. *Journal of Personality and Social Psychology*, 73, 345-368.

Binder, L.M., & Campbell, K.A. (2004). Medically unexplained symptoms and neuropsychological assessment. *Journal of Clinical and Experimental Neuropsychology*, 26, 369-392.

Dalton, P.H. & Jaén, C. (2010). Responses to odors in occupational environments. *Current Opinion in Allergy and Clinical Immunology*, 10, 127-132.

Eek, F., Karlson, B., Osterberg, K. & Ostergren, P. O. (2010). Factors associated with prospective development of environmental annoyance. *Journal of Psychosomatic Research*, 69, 9–15.

Gan, W.Q., Davies, H.W. & Demers, P.A. (2011) Exposure to occupational noise and cardiovascular disease in the United States: the National Health and Nutrition Examination Survey 1999–2004. *Occupational Environmental Medicine*, 68, 183-190.

Griffin, M.J., Bovenzi, M. & Nelson, C.M. (2003). Dose-response patterns for vibration-induced white finger. *Occupational Environmental Medicine*, 60, 16-26.

Levallois, P., Neutra, R., Lee, G., & Hristova, L. (2002). Study of self-reported hypersensitivity to electromagnetic fields in California. *Environmental Health Perspectives*, 110(Suppl 4), 619–623.
Levenstein S, Prantera C, Varvo V, Scribano ML, Berto E, Luzi C, & Andreoli, A. (1993). Development of the Perceived Stress Questionnaire: A new tool for psychosomatic research. *Journal Psychosomatic Research*, 37, 19-32.
Nordin, S., Karvala, K., Nyback, M. H., & Sainio, M. (2018). Prevalence of environmental annoyance in a Swedish and Finnish general population: Impact of everyday exposures on affect and behavior. *Journal of Environmental Psychology*, *56*, 84–90.
Nordin, S., Ljungberg, J., Claeson, A.-S., & Neely, G. (2013). Stress and odor sensitivity in persons with noise sensitivity. *Noise and Health*, *15*(64), 173.
Nordin, S., Millqvist, E., Löwhagen, O., & Bende, M. (2003). The Chemical Sensitivity Scale: Psychometric properties and comparison with the noise sensitivity scale. *Journal of Environmental Psychology*, *23*(4), 359–367.
Nordin, S., Millqvist, E., Löwhagen, O., & Bende, M. (2004). A short chemical sensitivity scale assessment of airway sensory hyperreactivity. *International Archives of Occupational and Environmental Health*, *77*(4), 249–254.
Nordin, S., Palmquist, E., Bende, M., & Millqvist, E. (2013). Normative data for the chemical sensitivity scale for sensory hyperreactivity: The Västerbotten environmental health study. *International Archives of Occupational and Environmental Health*, *86*(7), 749–753.
Nordin, S., Palmquist, E., Claeson, A.-S., & Stenberg, B. (2013). The environmental hypersensitivity symptom inventory: metric properties and normative data from a population-based study. *Archives of Public Health*, *71*(1), 18.
Nordin, S., Palmquist, E., & Claeson, A. S. (2013). Metric properties and normative data for brief noise and electromagnetic field sensitivity scales. *Scandinavian Journal of Public Health*, *41*(3), 293–301.
Overmier, J. B. (2002). Sensitization, conditioning, and learning: Can they help us understand somatization and disability? *Scandinavian Journal of Psychology*, *43*(2), 105–112.
Palmquist, E., Claeson, A. S., Neely, G., Stenberg, B., & Nordin, S. (2014). Overlap in prevalence between various types of environmental intolerance. *International Journal of Hygiene and Environmental Health*, *217*(4–5), 427–434.
Sparks, P.J. (2000). Idiopathic environmental intolerances: overview. *Occupational Medicine*, 15, 497–510.
Weinstein, N. D. (1978). Individual differences in reactions to noise: a longitudinal study in a college dormitory. *The Journal of Applied Psychology*, *63*(4), 458–466.

TIME PERCEPTION OF THE VERY DISTANT FUTURE

Hans-Rüdiger Pfister
Institute of Experimental Industrial Psychology (LüneLab), Leuphana University, 21335 Lüneburg, Germany
Gisela Böhm
DICE-Lab, Department of Psychosocial Science, Faculty of Psychology, University of Bergen, 5015 Bergen, Norway
pfister@uni-lueneburg.de, gisela.boehm@uib.no

Abstract

We examine how people perceive and represent the temporal structure of future events, using events occurring in the very distant future. While there is plenty of research about the perception of experienced duration of time intervals in the range of up to a few seconds, little is known about how people represent anticipated future events in the range of years, decades, or even centuries. Thinking about distant future events plays an important role in social and political decision making, for example, in discussions about how to deal with climate change, the consequences of which affect future generations. Employing a paired comparison approach, we asked participants to judge the subjective duration of the time interval between two future events. Intervals ranged from a few years to more than two hundred years. Based on these judgments, we use nonmetric multidimensional scaling to reconstruct the mental representation of the temporal structure of these events. Results suggest that the transformation of a symbolic description of future time points into a mental representation of time intervals has a logarithmic form, whereas the transformation of the mental representation into values of a response scale is linear.

Decisions are usually made after considering future consequences; the time when these consequences will occur has been shown to influence people's preferences. This problem has been studied under the topic of *time preferences* (Frederick, Loewenstein, & O'Donoghue, 2003). In this research, *time* per se has been construed as objective or physical time. The prevalent research question has been how the utility of a particular outcome is weighted according to how distant in the future the outcome will materialize, with *distance* measured in physical time units (days, years, etc.). With increasing distance, the utility of outcomes is usually *discounted* relative to immediate outcomes (Soman et al., 2005). Only recently has the question been addressed of how future time per se is perceived when making decisions (Han & Takahashi, 2012; Zauberman, Kim, Malkoc, & Bettman, 2009). In this study we examine the relationships between future time intervals, the mental representation of these intervals, and responses about the duration of these intervals.

Future Time

Whereas research on time discounting, originating in the economic sciences, has implicitly been based on physical time, the literature on time perception in cognitive psychology and psychophysics has long been aware of the fact that subjective time is not identical to physical time (Fraisse, 1984; Matthews & Meck, 2016). However, the majority of studies has examined the psychophysics of currently *experienced time* or *duration*, usually in the range of seconds or minutes (Grondin, 2010). There is some research on *past time* and *retrospective experience of duration* (Redelmeier, Katz, & Kahneman, 2003), but very little on *future time*.

Several issues complicate the study of future time. First, future time cannot be experienced (by definition). As we cannot present varying amounts of future time as a *physical* stimulus, classical psychophysical methods cannot be applied directly. Second, the perception of future time is always attached to particular events, for example, when an event will occur or how long the interval between two future events will be. This is especially apparent when studying time preferences, where the focus is on the valence of an *outcome* and how the outcome's utility changes as a function of time and valence. Third, presenting future time implies the use of a symbolic *number format* (e.g. "in 5 years"). How numbers are mapped onto subjective perceptions of quantities is still not fully understood (Dotan & Dehaene, 2016). If numbers are perceived in a non-linear way, any amounts of 'something' represented by numerical symbols may subjectively be distorted due to the numerical format. Findings from Schley and Peters (2014) suggest that the concave form of subjective utility functions may partially be due to a logarithmic transformation of the number line.

We assume the following:

- Given two future events e_1 and e_2, occurring at time t_1 and t_2, respectively (with $t_1 < t_2$), a *symbolic description* σ_1 and σ_2 represents these time points in some symbolic/numerical format (e.g.: "e_2 will occur 10 years after e_1"). For one event, the time point might be close to t_0, the immediate present.
- The symbolic description σ is perceived and mapped onto a mental representation of distances among the events. For a set of events $\{e_t\}$, the representation of all pairwise mental distances will be denoted as Ψ. We assume that Ψ can be modeled as a k-dimensional *metric space* (Shepard, 1987).
- Given the mental representation $\Psi(e_1,e_2)$, people can map the mental distance onto an appropriate *response scale* $R(e_1,e_2)$.
- Assuming a monotone relationship between Ψ and R, Ψ can be reconstructed from R using *non-metric Multidimensional Scaling* (nMDS). Given a suitable choice of k (dimensionality of Ψ) and r (metric of Ψ), the derived distances $d(e_1,e_2)$ from the nMDS configuration correspond to $\psi(e_1,e_2)$ (Nosofsky, 1992; Shepard, 1987).
- We presume three structures and two mappings: $\sigma(e_1,e_2) \rightarrow \psi(e_1,e_2) \rightarrow R(e_1,e_2)$. σ is the presentation as constructed by the experimenter, R are measurements on a response scale, and Ψ is estimated via nMDS.

Method

We conducted a computer-based experiment, presenting a series of pairs of future events that were described as occurring at two different future points in time. A sample of 22 participants (psychology students who participated as part of their course work) judged for all pairs of events the subjective duration of the interval between the pair.

Material

The fictional event to occur in the future was “The sea level will rise by several meters and will flood the coasts of the earth” (translation, all texts were presented in German). Eight future time points were used: 5, 30, 60, 90, 120, 150, 180, 210 years. Thus, 28 pairs of future events were presented, for example: “The sea level will rise by several meters and will flood the coasts of the earth in 30 years – or – The sea level will rise by several meters and will flood the coasts of the earth in 150 years.”

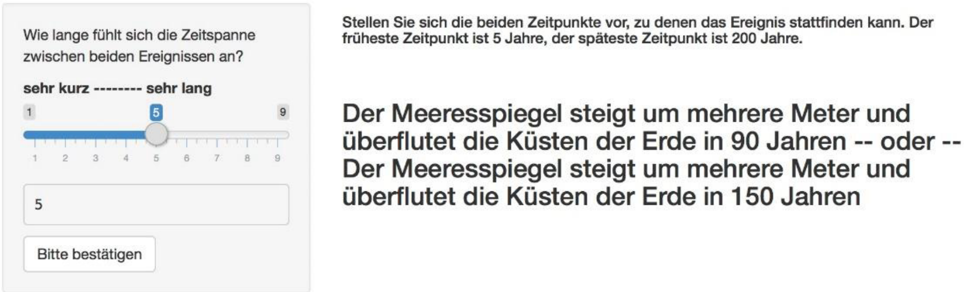


Fig. 1. Screen shot from the experiment showing the 9-point measurement scale on the left and the textual description on the right.

Procedure

The experiment was run individually on a computer, administered by an experimenter. Participants first performed a practice trial consisting of six pairs of events. The main experiment started with an instruction that emphasized the task to judge the subjective duration of the time interval between two future events. Then all 28 event pairs were presented in random order; the order of the two events within a pair was also randomized.

For each pair of events, participants judged the subjective duration of the interval on a 9-point scale anchored with ‘1 = very short’ and ‘9 = very long’. Participants adjusted the slider of a visual scale shown adjacent to the textual description of the events (Fig. 1). The experiment took approximately 15 minutes.

Results

For each participant i, the response matrix of judged durations R_i for all 28 pairs of 8 time points was obtained. All analyses are based on the average matrix R. To construct the mental representation Ψ, the R matrix was submitted to non-metric MDS (De Leeuw & Mair, 2009). Figure 2 depicts the two-dimensional configuration of future time points (Stress-1 = 0.025). It can be seen that dimension 1 represents a time dimension, whereas dimension 2 discriminates extreme from intermediate time points.

From the vector of the presented time points, the matrix D of *objective distances* was computed. Figure 3 shows the mapping of D on Ψ, that is, the *mental distances* derived from the nMDS as a function Ψ(D) of the objective distances.

We computed two regressions of Ψ on D: a non-linear logarithmic regression ($\psi = A \cdot \ln(d) + B$) and a linear regression. The logarithmic regression (solid line in

Fig. 3) yields a correlation between fitted and observed data of $r = 0.96$ and an $AIC = -40.66$, and the linear regression (dashed line in Fig. 3) yields a correlation of $r = 0.92$ and an $AIC = -19.89$, suggesting a better fit of the logarithmic function.

Figure 4 shows the mapping of Ψ on R, that is, the transformation of mental distances to corresponding values of the response scale. For this mapping, fitting a logarithmic function yields a correlation of $r = 0.95$ with an $AIC = 44.63$, and fitting a linear function yields a correlation of $r = 0.97$ and an $AIC = 25.12$, suggesting a better fit of the linear function.

Figure 5 shows the direct mapping of D on R. The logarithmic function with $r = 0.98$ and an $AIC = 21.92$ provides a better fit than the linear function with $r = 0.95$ and an $AIC = 46.48$.

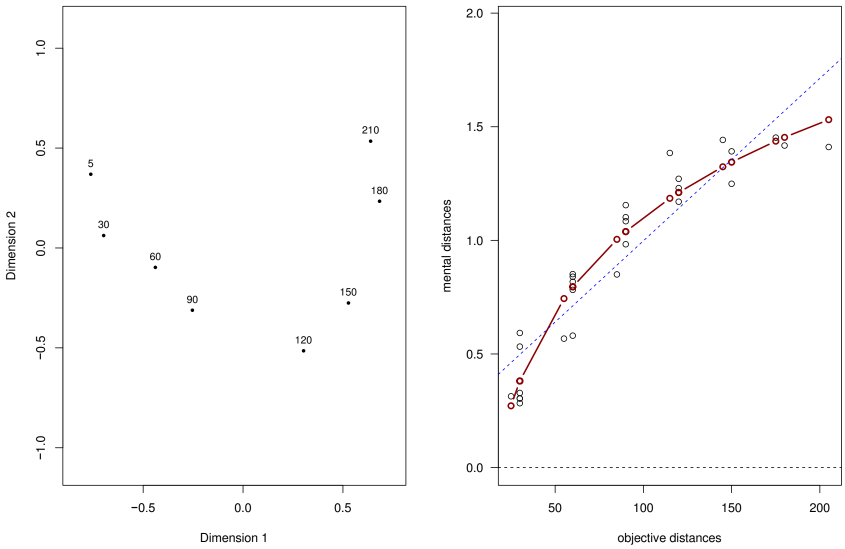


Fig. 2 (left) and Fig. 3 (right). Fig 2. shows the two-dimensional configuration from a nMDS of subjective responses. Fig 3. shows the mental distances as a function of the objective distances (solid line = logarithmic fit, dashed line = linear fit).

Discussion

Judging the duration of future time intervals requires two transformations: The transformation of a symbolic presentation of objective time intervals to a mental representation, and the transformation of mental distances to a response scale. Our data suggest that the first transformation is logarithmic ($D \rightarrow \psi$), and the second is linear ($\psi \rightarrow R$), yielding an overall logarithmic transformation (Fig. 3 to Fig. 5). This may depend on the particular way the symbolic times are presented, as well as on the type of response scale measuring subjective time durations.

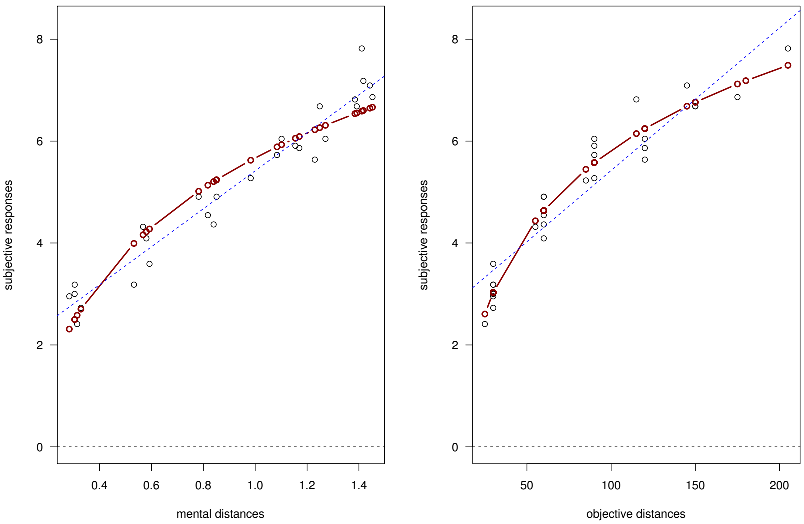


Fig. 4 (left) and Fig. 5 (right). Fig 4. shows the responses as a function of the mental distances. Fig. 5 shows the responses as a function of objective distances (solid line = logarithmic fit, dashed line = linear fit).

Acknowledgements

We thank Anita Wieczorek and Sarah Stritzke for collecting the data.

References

De Leeuw, J., & Mair, P. (2009). Multidimensional scaling using majorization: SMACOF in R. *Journal of Statistical Software, 31*(3), 1-30.

Dotan, D., & Dehaene, S. (2016). On the origins of logarithmic number-to-position mapping. *Psychological Review, 123*(6), 637-666. doi:dx.doi.org/10.1037/rev0000038

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

Frederick, S., Loewenstein, G., & O'Donoghue, T. (2003). Time discounting and time preference: A critical review. In G. Loewenstein, D. Read, & R. Baumeister (Eds.), *Time and decision: Economic and psychological perspectives on intertemporal choice* (pp. 13-86). New York: Russell Sage Foundation.

Grondin, S. (2010). Timing and time perception: a review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception, & Psychophysics, 72*(3), 561-582. doi:10.3758/APP.72.3.561

Han, R., & Takahashi, T. (2012). Psychophysics of time perception and valuation in temporal discounting of gain and loss. *Physica A: Statistical Mechanics and its Applications, 391*(24), 6568-6576. doi:10.1016/j.physa.2012.07.012

Matthews, W. J., & Meck, W. H. (2016). Temporal cognition: Connecting subjective time to perception, attention, and memory. *Psychological Bulletin, 142*(8), 865-907. doi:10.1037/bul0000045

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology, 43*, 25-53.

Redelmeier, D. A., Katz, J., & Kahneman, D. (2003). Memories of colonoscopy: A randomized trial. *Pain, 104*(1-2), 187-194.

Schley, D. R., & Peters, E. (2014). Assessing "economic value": Symbolic number-mappings predict risky and riskless valuations. *Psychological Science, 25*(3), 753-761. doi:0.1177/0956797613515485

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237*, 1317-1323.

Soman, D., Ainslie, G., Frederick, S., Li, X., Lynch, J., Moreau, P., . . . Zauberman, G. (2005). The psychology of intertemporal discounting: why are distant events valued differently from proximal ones? *Marketing Letters, 16*(3/4), 347-360.

Zauberman, G., Kim, B. K., Malkoc, S. A., & Bettman, J. R. (2009). Discounting time and time discounting: Subjective time perception and intertemporal preference. *Journal of Marketing Research, 45*(4), 543-556. doi:10.1016/j.obhdp.2005.01.001

CHANGE IN ODOR ASSESSMENT THROUGH COGNITIVE EMBEDDING

Vanessa Thiele, Malte Wiethege, and Kirsten Sucker
Institute for Prevention and Occupational Medicine of the German Social Accident Insurance. Institute of Ruhr University Bochum, D-44789 Bochum, Germany

Several studies have shown that manipulation of expectations and beliefs about health-risks from airborne chemical exposure can alter chemosensory perception ratings, symptom reports, and even cognitive performance measures. Furthermore, ratings of odor pleasantness vary as a function of labels assigned to the odor. Typically, unidimensional rating scales like category or visual analog scales are used in such experiments. To the best of our knowledge, manipulation of odor pleasantness assessed with the "Polarity Profile" method has not been done before.

Two odorants, benzaldehyde with an almond-like, pleasant odor, and naphthalene with an unpleasant odor reminding of mothballs, were evaluated using the Polarity Profile method, i.e. a list of 29 semantic differential scale adjectives. Information bias about the nature of the odorant varied (pos. - natural extract used in aromatherapy; neg. - industrial solvent used in the chemical industry) as well as presentation order. Hence, participants were assigned to one of four groups:

(1) benzaldehyde (pos.) – naphthalene (neg.),
(2) benzaldehyde (neg.) – naphthalene (pos.),
(3) naphthalene (neg.) – benzaldehyde (pos.),
(4) naphthalene (pos.) – benzaldehyde (neg.).

In a pre-study, only participants in group (4), with benzaldehyde on second place and with the negative information bias, could be influenced. They rated the pleasant almond-like odor as more unpleasant than the participants in the other three groups.

The results show, that both the information and the order of presentation affect the ratings. In the main study, a larger sample with a wider age range and equal distribution of gender will be realized.

THE SHAPE-WEIGHT ILLUSION AND THE BRIGHTNESS-WEIGHT ILLUSION

Michele Vicovaro, Katia Ruta, and Giulio Vidotto
Department of General Psychology, University of Padova
michele.vicovaro@unipd.it

Physical weight is only one of several variables that may affect weight perception. For instance, when two objects of the same physical weight but different sizes are lifted, the smaller object usually feels heavier than the larger object (i.e., the size-weight illusion). A unified explanation of weight illusions is provided by the so-called *expectation model*, according to which perceived weight results from the contrast between the sensory information about the actual weight of an object and the cognitive information about its expected weight. Because of that contrast, whenever two objects of the same physical weight are expected to have different weights, the object that is expected to be heavier is perceived to be lighter than the other object.

In the current study, we focused on two relatively unexplored weight illusions, namely the shape-weight illusion and the brightness-weight illusion. Specifically, we sought to explore whether these two illusions can be accounted for by the expectation model. The stimuli were 18 three-dimensional plastic objects of identical weight (i.e., 250 g), resulting from the orthogonal combination of three volumes (216, 729, 1728 cc), three shapes (cube, sphere, and tetrahedron), and two colors (black and white). A large 4580 cc, 450 g red cylinder served as the standard stimulus. In the *perceived size condition*, participants provided a visual volume estimate of the stimuli; in the *expected weight condition*, participants provided a visual estimate of the apparent weight of the stimuli (i.e., touching the stimuli was not allowed); in the *perceived weight condition*, participants were allowed to lift the stimuli by using a string that was attached to the top. In each of these conditions participants were asked to provide an estimate of the target dimension (i.e., perceived size, expected weight, or perceived weight), relatively to the standard that was assigned value 100. Thirty-four participants took part in the expected and the perceived weight conditions, and a subset of 20 of these 34 participants also took part in the perceived volume condition.

As for the *shape* factor, the results showed that tetrahedrons were perceived to be larger, and expected to be heavier, than spheres and cubes, especially at the largest level of the volume factor. Spheres were perceived to be larger and expected to be heavier than cubes, though not at a statistically significant level. Consistently with the predictions from the expectation model, tetrahedrons were perceived to be lighter than spheres and cubes, but, inconsistently with the predictions from the model, spheres were perceived to be heavier – rather than lighter – as compared with cubes. As for the *brightness* factor, the results showed that – at the largest level of the volume factor – white stimuli were perceived to be larger than black stimuli; at the lowest level of the volume factor, black stimuli were expected to be heavier than white stimuli; intriguingly, black stimuli were perceived to be heavier than white stimuli at the lowest level of the volume factor, whereas the opposite was true at the largest level of the volume factor. Neither of the latter two results is consistent with the predictions from the expectation model.

Overall, the results suggest that the influence of shape and brightness on the perceived volume, expected weight, and perceived weight of objects is strongly mediated by the objects' physical size. The results are only partially consistent with the predictions from the expectation model.

Author Index

FECHNER DAY 2018
LÜNEBURG