The 37th Annual Meeting of the

International Society for Psychophysics

Conference Proceedings

Fechner Day 2021

October 19-22, 2021

Published by the International Society for Psychophysics, 2021.

Cover Image: Markus Spiske

Cover design: Jordan Richard Schoenherr

Title: *Proceedings of the 37th Annual Meeting of the International Society for Psychophysics Fechner Day 2021*

Edited by Jordan Richard Schoenherr, Timothy Hubbard, and Kazuo Ueda for the International Society for Psychophysics.

Table of Contents

01010111 01100101 01101100 01100011 01101111 01101101 01100101 00100000
01001101 01100101 01110011 01110011 01100001 01100111 01100101

WELCOME

Welcome to the 37th Annual Conference of the International Society for Psychophysics, Fechner Day 2021!

This year's annual meeting marks our second year as an online event. As we begin to see the tail of the pandemic close, the Executive Committee is optimistic that 2022 will allow for a more traditional meeting. Nevertheless, I was pleased once again to see that the calibre of submissions was on a par with previous years, presenting the usual range of research using a variety of methods across many modalities. Gustav Theodore Fechner would surely approve of our continued mix of eclectic approaches to the study of sensation and perception.

Coordinating an international meeting across all time zones brings with it many challenges, as does holding during the regular academic year. We are appreciative of the time and effort that our members commit every year to the study and promotion of psychophysical research. At this year's conference, we have papers surveying the deep historical roots of physiological psychology, contemporary theoretical frameworks that attempt to capture the growing body of psychophysics, empirical works that push foundational research in sensation and perception forward.

As with other years, there are many people deserving of credit for their efforts. As always, Strongway and Wolfgang provided the technical and financial infrastructure for our meetings. Natalia Postnova proposed, and created, our new Google Group, helping to address many of the challenges we faced due to the overexuberance of spam filters. Once more, Tim Hubbard helped this year by reviewing submissions.

Finally, I cannot forget to thank our President, Kazuo Ueda and the rest of the Executive Committee for their assistance in organizing.

Hoping to see you all again in 2022!

Jordan Richard Schoenherr, PhD

Department of Psychology, Concordia University

Department of Psychology / Institute for Data Science, Carleton University

Committee Membership

Organizing / Executive Committee

Wolfgang Ellermeier

Leah Fostick

Natalia Postnova

Jordan Richard Schoenherr

Kazuo Ueda

Program Committee

Jordan Richard Schoenherr

Timothy Hubbard

Proceedings Committee

Jordan Richard Schoenherr, Chair

Timothy Hubbard

Kazuo Ueda

OKTOBER 22 – Fechner's Nose

A Cavern near Elora

Thrice the brinded cat hath mew'd
Thrice and once the hedge-pig whin'd
Harper cries: 'Tis time, 'tis time.

Round about the cauldron go
In the poison'd entrails throw
Prized fillet of fenny snake
In the cauldron boil and bake
Eye of newt and toe of frog
Wool of bat and tongue of dog
Adder's fork and blind worm's sting
Lizard's leg and Howley's sting
For a charm of powerful trouble
Like a hell broth boil and bubble.

Scale of dragon, ear of wolf
Witches mummy, maw and gulf
Psychics beard and tooth of shark
Roots of Hemlock and Herbart
Eyes of Newton and nose of Weber
Cast into steaming vapor
Swirling vortex of vague suggestions
Roil and burble beneath perception
Enrage the fire primeval mystiks
Bring to birth our Psychphysiks.

Oh, well done! I commend your pains
And everyone shall share i' the gains
Now around the cauldron ring
Novices fated soon to sing
Praises of the Elements
Echoed by experiments

Hark!
Anon!

By the prickling of my thumbs
Gustav Fechner this way comes.

Stephen Link
[after William Shakespeare]

SCHEDULE

<u>Symposia Day 1 (Tues Oct 19)</u>

Start Time: 08:00 EST (Americas):: 14:00 CEST (Europe/MidEast):: 21:00 JST (Japan)

Time (EST) **Start** **Stop**	**TITLE**	**AUTHORSHIP**
07:50 *08:00*	**Pre-Session Mingle**	
08:00 *08:03*	**Welcome Message and Session Introduction**	Jordan Richard Schoenherr
08:03 *08:05*	**Introduction**	Kazuo Ueda
08:05 *08:35*	**Helmholtz and the Impact of the Stereoscope on Visual Science in Germany**	Nick Wade
08:35 *08:40*	**Extended Question Period**	
08:40 *08:45*	**Break**	
08:45 *09:00*	**Higher-Order and Medial Psychophysics: A New Typology**	Tim L. Hubbard
09:00 *09:15*	**Feathers, Metal and Weight Expectations**	Helen E. Ross
09:15 *09:30*	**Fechner's Methods for Measuring Differential Sensitivity:** **Do the Results Correlate for Visual Tests of Line Length?**	Nan Yang, […] Gordon Waddington
09:30 *09:45*	**Evaluating Binary Decisions: Models, Predictors, and Collaboration**	Diana Kornbrot
09:40 *09:45*	**Closing session**	

<u>Symposia Day 2 (Weds Oct 20)</u>

Start Time: 07:00 EST (Americas):: 13:00 CEST (Europe/MidEast):: 20:00 JST (Japan)

Time (EST) **Start** **Stop**	**TITLE**	**AUTHORSHIP**
07:00 *07:05*	**Greetings**	
07:05 *07:20*	**Smell Infant Psychophysics: Methodological adaptation for COVID-19 Pandemic**	Rosana Tristão … Thomas Lachmann
07:20 *07:35*	**Modality Specificity of Amodal Completion**	Hiroshige Takeichi, Keito Taniguchi, & Hiroaki Shigemasu
07:40 *07:55*	**Spatiotemporal Brain Mechanism of Auditory and Tactile Time Shrinking**	Takako Mitsudo, … Shozo Tobimatsu
07:55 *08:10*	**The Effect of Stimulus Duration on the Perceived Congruency of Stimuli Consisting of a Gabor Patch and an AM- or FM-tone**.	Natalia Postnova & Gerard B. Remijn
08:10 *08:15*	**Break**	
08:15 *08:30*	**Checkerboard Speech: A New Experimental Paradigm for Investigating Speech Perception**	Kazuo Ueda, Riina Kawakami, & Hiroshige Takeichi
08:30 *08:45*	**Your Money or Your Life?!: Social (Psychological) Distance and Temporal Discounting**	Jordan Richard Schoenherr
08:45 *08:50*	**Closing Session**	

Symposia 3 (Thurs Oct 21)

Start Time: 08:00 EST (Americas):: 14:00 CEST (Europe/MidEast):: 21:00 JST (Japan)

Time (EST) Start Stop	TITLE	AUTHORSHIP
08:00 *08:05*	**Greetings**	
08:05 *08:20*	**Temporal and Frequency Resolution Needed for Auditory Communication: Comparison Between Young and Senior Listeners Utilizing Mosaic Speech**	Yoshitaka Nakajima … Gerard B. Remijn
08:20 *08:35*	**Acoustic Correlates of English Consonant-Vowel-Consonant (CVC) Words Obtained with Multivariate Analysis**	Yixin Zhang … Gerard B. Remijn
08:35 *08:50*	**Psychology Teaching by Using Volumetric Video Based AR and Avatar AR Technology is More Attractive Than Those by Face to Face and Zoom**	Xuanru Guo … Takeharu Seno
08:50 *09:05*	**The Role of Afterimages in Motion Induced Blindness**	Sofia C. Lombardo … Wm Wren Stine
09:15 *09:25*	**Break**	
09:25 *09:40*	**The Effect of Different Spatial Sound Distributions on Cognitive Load in Audiovisual Perception Tasks**	Catarina Mendonça
09:40 *09:55*	**The Dynamics of Perceptual Filling-in and Motion Induced Blindness during Binocular Rivalry: Exploring the Common Mechanism Hypothesis**	Rebecca White … Wm Wren Stine
10:10 *10:25*	**Global Temporal Context and Sensory Modality Effects on the Discrimination of Short Rhythmic Sequences**	Hugo Fitzbak-Fortin … Simon Grondin
10:25 *10:30*	**The Gain and Loss from Inter-Channel Interactions in the Parallel Systems**	Yanjin Liu & James T. Townsend
10:30 *10:35*	**Closing Session**	

Symposia Day 4 (Fri Oct 22)

Start Time: 07:00 EST (Americas):: 13:00 CEST (Europe/MidEast):: 20:00 JST (Japan)

Time (EST) Start Stop	TITLE	AUTHORSHIP
07:00 *07:05*	**Greetings**	
07:05 *07:20*	**Vection Depends on the Luminance Contrast, Average Luminance and the Spatial Frequency of the Stimulus**	Xuanru Guo … Stephen Palmisano
07:20 *07:35*	**Eye Movement Abnormalities among Patients with Schizophrenia**	Alexandra Wolf, Kazuo Ueda, & Yoji Hiran
07:40 *07:55*	**On the Density of Faces in Face Space: Explanation in Terms of Infinite Dimensional Riemannian Geometry**	James T. Townsend, Cheng-Ta Yang, & Hao-Lun Fu
07:55 *08:10*	**Mentalarcheology**	Stephen Link
08:10 *08:25*	**Numerosity Perception: Investigating the Temporal Dynamics using Non-Symbolic Stimuli**	Aslı Bahar İnan and Aslı Kılıç
08:25 *08:40*	**How do We Perceive Masked Faces during COVID 19 Pandemic?**	Evrim Gülbetekin … Elif Ezgi Kaplan
08:45 *08:40*	**Closing Session**	
08:40 *09:10*	**BUSINESS MEETING**	

Paper Index

Abstracts and Papers

:: Listed Alphabetically by First Author ::

Global Temporal Context and Sensory Modality Effects on the Discrimination of Short Rhythmic Sequences

Hugo Fitzback-Fortin, Esteban Mendoza-Duran and Simon Grondin
École de Psychologie, Université Laval, Québec, Canada

Abstract—The present study aimed to observe if the temporal context of a sample of different sequence of intervals would exert the same effect on individuals depending on the sensory modality of the stimuli. To that end, an experimental paradigm similar to Jones and McAuley (2005) was used to further explore the propensity of time distortions to happen while manipulating modality and temporal context. The results show not only that temporal context influences time perception when the modality of the stimuli is auditory and visual or a combination of both, but even more so, there is a modality-specific effect on the perception of time. Both effects exert their influences on time perception without trumping one another and interacting together only for constant error.

Keywords— Rhythm, time perception, sensory modality

I.INTRODUCTION

Rhythm is a facet of time whose scope includes the limits of music theory and the perception of time. What is more, it is as much the domain of language, with its implications in movement and action (Large & Jones, 1999). Rhythm can also be summed up as a temporal pattern: a set of time intervals delimited by perceptible events giving the impression of a periodicity or a beat (McAuley, 1995). Therefore, a rhythm can emerge when the perceived stimuli integrate on a timeline, according to the relationship maintained between the durations of the stimuli and that of the silence that separates them (Jones & Boltz, 1989).

Considering the above, examples abound with inherent rhythms from the environment. For example, just consider cyclical events like day and night, the seasons, or just the arrival time of a bus. In addition, biological phenomena in animals and humans exhibit rhythmic characteristics that closely relate to time (Boltz, 1994). Indeed, language and communication, the cycle of sleep and wakefulness, as well as locomotion are excellent examples of rhythmic phenomena specific to living things. In addition, several other internal biological phenomena unfold over time periodically (Nobre & van Ede, 2017). In such a rich temporal environment where the rhythmic characteristics of stimuli are as numerous as the elements which conceal them, what main theories can account for the perception of time and rhythm adapted to this environment?

II.THEORITICAL MODELS

Scalar expectancy theory. Despite the inexistence of a specific sensory organ or cerebral structures dedicated to the perception of time, one of the most frequently used models suggests that physical time is perceived based on an internal clock. According to Gibbon's Scalar Expectancy Theory (SET; 1977), this internal clock is a pacemaker-counter device. In this model, a counter stores pulses that are emitted by a pacemaker, and it is the accumulated number of pulses that determines the estimated time (Church et al., 1984). Some authors posit that the counter is also equipped with several modules that relate to different sensory modalities and that count the pulsations differently (Rousseau & Rousseau, 1996).

SET applies to many situations. It accounts for the effects of different factors that can affect the perception of time. It simply and flexibly integrates elements (variations of arousal or attention to time) into its explanatory structure. However, it cannot interpret certain phenomena that are more dynamic than those normally encountered in experimental tasks (McAuley & Jones, 2003; Jones et al., 2002). SET is not well suited to account for some effects related to dynamic contexts.

Dynamic attending theory. Originally developed by Jones (1976), DAT adequately applies to the dynamism inherent in the concept of rhythm and aims to specifically explain the processes involved in paying attention to the present (Large & Jones, 1999). This theory explains perceived duration according to an approach of relative time. In this perspective, time is not only perceived as a series of absolute points and intervals, but rather in relation to other time periods unfolding in the environment. They form a global structure composed of time periods determined by the relative velocity of events and the flow of the structure they compose (Jones & Boltz, 1989). Jones and Boltz (1989) maintain that subjective time depends on the rhythmicity between the different elements currently perceived. This rhythmicity can be projected out of its current perception and thus extended into the future, which then allows for the prediction of things to come. Finally, the acuity of judgments about time depends on the temporal coherence of the perceived events and the ability to synchronize the attentional rhythm with the appropriate level of rhythm that the environment can offer.

Due to the chaotic nature and noise of external rhythms, perceptive organisms deal with this lack of absolute coherence by integrating perceivable rhythms and merging them together into a global impression. Temporal context is thus defined as the composition of multiple rhythms nesting in one another in an hierarchical fashion (Jones & Boltz, 1989). This indicates, as a global impression, the expected appearance of a stimulus by its own rhythm and that of the others inscribed within the context (Barnes & Jones, 2000). Accordingly, while perceiving time within a temporal context, the duration to be judged is often categorized in accordance with the global value of durations presented, rather than the real duration (Barnes & Jones, 2000).

This temporal context effect is captured by DAT as an externally induced internal rhythm, following the conception of Fraisse (1963) of a biological inner rhythm named a *preferred period*. Contrary to its first definition, Jones and McAuley (2005) define the preferred period as an adapting internal rhythm, the variations of which are caused by oscillating inner processes synchronizing themselves with rhythms emitted by external sources, mainly the temporal context. Therefore, the preferred period is not fixed and varies according to the global context. The preferred period can be « moved around » when the temporal context changes and can be indexed as a measure of its effect.

III. SENSES AND RHYTHM

The efficiency of time perception varies according to the sensory modalities that mark intervals (Grondin, 2010, 2020). For example, the interval needed between two visual stimuli to distinguish two stimuli instead of only one has to be longer than that between two sounds (Jones, 2014). Furthermore, Levitin et al. (2000) show that when using intermodal conditions combining hearing and sight, a certain lag is required between a sound and a flash for them to be considered as simultaneous. Differences between sensory modalities are especially observed when time intervals are to be judged. When comparing the perceived duration of two time intervals of the same duration, one delimited by sounds and the other by flashes, the one composed of sounds will be perceived as being longer (Rammsayer, Bortern, & Troche, 2015). A time interval marked by sounds is also recognized as being easier to discriminate than a time interval marked by flashes (Azari et al., 2020; Grondin, 1993).

Secondly, differences in sensory modalities are also observed during an anchoring effect in tasks of judging durations. The anchoring effect consists in the presentation of a first stimulus whose duration differs greatly from the following ones. The duration of this first stimulus persists when judging the sequence of intervals as a whole and influences the judgment in such a way as to overestimate or underestimate the duration of the sequence. Behar and Bevan (1961) report that the use of a sound stimulus as an anchor when the rest of the stimuli are lights causes a weaker anchoring effect compared to a visual modality anchor. The reverse is also observed, a visual anchoring in a sequence of sounds having a reduced effect compared to an anchoring of the same modality. However, the reduction generated with visual anchoring is smaller than with sound anchoring under intermodal conditions. In short, the modality of the stimuli making up these intervals influences the judgments of durations.

Moreover, Gamache and Grondin (2010) also report greater acuity in an interval reproduction task when two modalities are involved. Indeed, the reproduction of an interval presented with overlapping auditory and visual modalities is more successful than when intervals are presented using only one of those modalities (Gamache & Grondin, 2010). Although these experiments are not specific to rhythm, it appears that the involvement of several senses in a time perception task improves performance through the rhythmic synchronization of the senses between them.

IV.OBJECTIVES

The present study aims to observe if global temporal effects and modality effects interact or maintain themselves when they are both engaged in a sequence paradigm.

To these ends, the experimental paradigm used in the study will consist in the presentation of a sequence of stimuli spaced by the same duration, followed by a pair of stimuli defined as a comparison interval, which are spaced by a duration that is either shorter, longer, or identical to the base duration of the sequence (Fig.1). The sequence is composed of three intervals for which the first two must be ignored and the third memorized. For each trial following this presentation, differences in the performance of the participants due to the variation of global distributional characteristics of a session (mean interval, range, and number of different intervals) and the sensory modalities engaged should be observed.

V.METHOD

Participants. Twenty-three participants, from 18 to 30 years old, were recruited for the experiment. All of them are University students contacted by email through the Laval University messaging system. At the end of recruitment, two participants were excluded for having a mean accuracy that was lower than chance level and three others for failing to complete the task. A total of 18 participants (6 male and 12 female; *M* age = 25) completed the task and had an accuracy level above chance.

Material. The task was conducted on a fractal design computer with a square screen using the experimentation software E-prime 3.0. The computer screen had a 25-in diagonal and a refresh rate of 240 Hz. The sounds used in the experiment were transmitted through a pair of Logitech multimedia speakers Z200. As is the case in Jones and McAuley (2005), the 440-Hz, 60-ms sounds had an intensity varying from 60 to 70 dB and were generated with an online tone generator (https://www.szynalski.com/tone-generator/). In the visual condition, an 8-cm diameter black circle appearing on a grey screen was also presented for 60 ms.

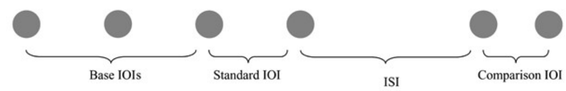


Figure 1. Sequence of intervals composed of a context inducing base interval, a standard interval to remember followed by a comparison interval to be compared with the standard.

Procedure. Participants completed eight experimental sessions, with four blocks of 36 or 60 trials (see below) per session. Each session corresponded to one of eight conditions: four modality conditions delimiting time intervals × two mean interval lengths. In two of the four modality conditions, all intervals were either delimited by sounds or delimited by circles; and in two other conditions, sounds occurred before circles, or circles before sounds. The two mean intervals lasted 400 or 600 ms.

In a session, two blocks presented five different time intervals, which ranged from 200 ms (narrow) or 400 ms (wide); the other two blocks presented three time intervals with the same ranges. No matter the number of time intervals, an interval corresponding to the mean would always be presented. Each standard duration in a block would appear 12 times. Consequently, depending on the number of different time intervals, the number of trials per block is either 36 or 60 trials. In total, a session of four blocks would have 192 trials. Therefore, to complete all sessions, participants went through 1536 trials and judged each interval 96 times. Completing a session took 45 to 90 minutes. Between each trial block, participants were given the opportunity to take a break, which lasted no longer than five minutes.

Participants completed the four blocks in a randomized order determined for each session. Experimental trials varied according to the length of the standard, the length of comparison intervals, and which sensory modality was being presented. The standard intervals had durations ranging from 200 ms to 800 ms. Comparison intervals were 12% or 24% shorter or longer than the standard intervals. The comparison intervals were in some cases identical to the standard. There was an equal number of short, long, and identical comparison intervals in each block. The interval between the occurrence of the stimulus (sound or circle) ending the standard interval and the occurrence of the stimulus (sound or circle) beginning the comparison interval was always twice the duration of the standard intervals. Finally, the range from the smallest time interval to the highest within a block was 200 ms in two blocks and 400 ms in the other two.

Overall, two mean standard intervals, two ranges of intervals, two different numbers of intervals and four modality combinations are crossed to obtain 32 experimental conditions. From there on, the experiment followed a 2 × 2 × 2 × 4 factorial design. Every participant was directed through each condition.

Data analyses. Two dependent variables are of interest: the probability of responding correctly (PC) and the constant error (CE). The PC is the average probability of a correct response over all comparison intervals within a condition. The CE is determined by subtracting each base interval from the average point of subjective equality (PSE) in each condition. Accordingly, a PSE is determined for each condition as in the method used by Jones and McAuley (2005).

To properly estimate the PSEs in a three-choice task, the probability of responding "long" to a comparison interval is subtracted from the probability of responding "short". This approach is adopted for each of the five comparison intervals and is referred to by Jones and McAuley (2005) as a *difference method*. A PSE, following this method, is defined as the standard intervals that produce a probability of responding "short" minus "long" equal to zero. Hence, to estimate the PSEs, the probability values are rescaled from the boundary of -1 to 1, to a scale of 0 to

1. As a result, psychometric curves could be approximated as cumulative normal distributions. These rescaled values were then transformed again into z coordinates to yield a linear psychometric function. The PSE, *per se*, is calculated by regressing the z coordinates over the comparison intervals for each base interval. The intersection of the regression line and the abscissa indicated the PSE.

For each dependent variable, PC, and CE absolute score, an ANOVA according to a 2 (context mean) × 2 (range) × 2 (number of IOI) × 4 (modality conditions), with repeated-measures on all factors, was conducted. Furthermore, to determine the preferred period of each sensory condition, eight regression lines, two per modality and one for each mean interval, were estimated from the CEs.

VI.RESULTS

Proportion of correct responses (PC). A linear regression was conducted to verify if the factors used were all significantly accounting for the variance of the PC. The stepwise regression showed that adding range and context to the first model, i.e. the model which includes only modality ($R^2 = 0.088$, $F(1, 574) = 55.421$, $p < 0.001$), provides a model explaining a significant amount of variance ($R^2 = 0.139$, $F(1,572) = 14.064$, $p < 0.001$) of the observed PC. At each entry, the number of IOI failed to attain a significance level of $p < 0.05$ ($p = 0.260$), and thus, this factor was excluded from the three regression models. Because of this linear regression, further analyses will not consider the number of IOI as relevant to the prediction of P, to the difference between PC scores, and to the differences observed between conditions for CE.

Subsequently, the repeated-measures ANOVA on PC, without taking the numbers of IOI into account, were significant for context, $F(1, 17) = 22.594$, $p < 0.001$, range, $F(1, 17) = 94.982$, $p < 0.001$, and range and modality effects, $F(2, 30) = 26.900$, $p < 0.001$. The PC are higher in the 600-ms mean session ($M = 54.9$, $SE = 1.92$) than in the 400-ms one ($M = 51.8$, $SE = 2.01$), and in the narrow-range condition ($M = 55.44$, $SE = 1.88$) than in the wide-range condition ($M = 51.21$, $SE = 2$).

Pairwise comparisons for the modality conditions showed that the auditory condition ($M = 60.71$, $SE = 2.84$) led to a significantly better performance than all other conditions ($p < 0.001$), that the visual condition ($M = 51.79$ $SE = 1.78$) did not lead to a significantly better performance than the visual-auditory condition ($M = 51.11$, $SE = 1.85$), nor the auditory-visual condition ($M = 49.66$, $SE = 1.72$).

Constant Error (CE). Following the analysis of the CE adopted by Jones and McAuley (2005), an individual function in each experimental condition was drawn, plotting the probability to respond "short" minus "long" on the y axis, as a function of the comparison intervals. This probability was transformed in Z scores, and a linear regression was conducted. The PSE was obtained by finding the value on the function corresponding to a Z score of 0. Then, on the basis of average PSE for each condition, CE scores were calculated by subtracting the base IOI from the PSE.

The goodness-of-fit for the psychometric functions was assessed by the R^2 statistic. R^2 was above 0.9 for 85.16% of the cases; for the remaining 14.84%, $R^2 > 0.7$. Averaged over participants, this yielded 32 CE scores per modality condition (total of 128). CE scores were always positive when the base IOI was faster than the mean session, and negative when slower than the mean. Moreover, the more remote the base IOI was from the mean, the greater was the magnitude of the CE.

A 2 (context mean) × 2 (range) × 4 (modality conditions) repeated-measures ANOVA was conducted on the absolute values of CE scores. The analysis revealed that the mean session effect was significantly lower at 400 ms ($M = 34.22$; $SE = 6.848$) than at 600 ms ($M = 43.90$; $SE = 9.45$), $F(1,7) = 5.63$, $p < 0.05$; the CE was also significantly higher with a wide range ($M = 44.94$; $SE = 10.15$) than with a narrow range ($M = 33.12$; $SE = 6.06$), $F(1,7) = 5.83$, $p < 0.05$; and the auditory ($M = 26.07$; $SE = 4.80$), visual ($M = 43.94$; $SE = 11.07$), auditory-visual ($M = 40.91$; $SE = 7.44$), and visual-auditory ($M = 45.19$; $SE = 11.17$) conditions also differed significantly, $F(3, 21) = 3.33$, $p < 0.05$.

In addition, the interaction between session mean and modality condition was found significant, $F(3,21) = 3.937$, $p = 0.022$. The simple effect tests revealed that CE scores differ significantly between modalities when the session mean is 400 ms, $F(3, 45) = 5.055$, $p = 0.004$ and when the session mean is 600 ms, $F(3, 45) = 3.640$, $p = 0.020$. With the 400-ms condition, the auditory condition differed significantly from the visual condition ($p = 0.011$), but not from the auditory-visual ($p = 0.070$) and visual-auditory conditions ($p = 1$). With the 600-ms condition, none of the modality conditions differed from each other.

Finally, the *preferred period* was obtained by regressing CE scores over base IOI and resolving the regression equation to calculate the base interval for a CE = 0. Thus, 16 regression lines were estimated based on the mean CE scores over participants: four per modality, each one representing the combination of the session mean interval of 600 ms and 400 ms with the range of 200 ms (narrow) and 400 ms (wide). All regression equations explained a significant amount of variance ($p < 0.05$) for their designated condition. Twenty five percent of the fits were above $R^2 = 0.9$, 31.25%, others were above $R^2 = 0.7$, and the smallest fit value obtained (for the narrow audio-visual condition within a 600-ms session mean) was $R^2 = 0.58$. For each condition, the 16 separate regression lines provided the location of P when CE = 0.

VII.CONCLUSION AND LIMITS

The present study shows that the pattern of over and under indexed by the absolute magnitude of CE scores is indicative of an internal representation of pace. Moreover, this preferred period is influenced by the distributional characteristics of several short rhythmic sequences and the sensory modality of the stimuli marking the time intervals. Thus, the global temporal context instills varying representations of rhythm that depend on the mean duration perceived, the range of durations observed, and the sensory systems at play. Mean duration and sensory modality seem connected, interacting with each other in the elaboration of the representation of pace.

However, the global distributional characteristics and the modality that were used may not give a complete portrait of how an internal representation of pace is obtained. For instance, Jones (2019) talks about the possibility of a *temporal niche,* a concept akin to the environmental niche of a living being. Multiple time imperatives in the form of events with stable durations tend to repeat themselves, and living creatures learn to adapt to these time constraints by entraining specifically to these time intervals. With repetition comes excellence, and they thus become adapted to perceiving durations more often encountered in their environment. Thus, as living beings, we are more accustomed to durations and paces encountered in our everyday life. Over extremely longer periods, entrainment occurs, and an internal pace is acquired through environmental needs. Way before a time experiment is executed in laboratory, participants have an inner sense of pace entrained to their respective time imperatives, independent of the distributional characteristics and modality of time markers.

This fact motivated the use of repeated measures in the experimental design of this study to prevent differences based on individual inner pace. Yet, it also gave rise to a limit of the present design. Indeed, by executing every condition, participants had to entrain themselves in a randomised order to different temporal contexts. It is possible that different pace representations can be obtained with different orders of presentation for a single participant, a possible influence that could not have happened to the same extent in Jones and McAuley (2005) who used a between-subjects design. Their participants entraining to a specific context, but not to others, they had a less variable representation of pace, although the possibility of an influence from a *temporal niche* could not be avoided.

Further studies on rhythm perception within the same framework as the one used in this study should then take into consideration that participants have an individual sense of pace acquired through their own experience. They should also study the influence of the order of presentation of different temporal contexts over participants representation of pace. Finally, a better understanding of the elaboration of a sense of pace could be gained by exploring if the effects of entrainment are time sensitive and vary depending on the delay between entrainment sessions. This could also pave the way for studies aiming to further define how a *temporal niche* is established and how subjective time is affected by it.

References

Azari, L., Mioni, G., Rousseau, R., & Grondin, S. (2020). An analysis of the processing of intra- and intermodal time intervals. *Attention, Perception, & Psychophysics*, *82*, 1473-1487. doi :10.3758/s13414-019-01900-7

Barnes, R., & Jones, M. R. (2000). Expectancy, attention, and time. *Cognitive Psychology, 41,* 254-311. doi: 10.1006/cogp.2000.0738

Behar, I., & Bevan, W. (1961). The perceived duration of auditory and visual intervals : Cross-modal comparison and interaction. *The American Journal of Psychology, 74,* 17-26. doi:10.2307/1419819

Boltz, M. G. (1994). Changes in internal tempo and effects on the learning and remembering of event durations. *Journal of Experimental Psychology, 20,*1154-1171. doi:10.1037/0278-7393.20.5.1154

Church, R. M., Meck W. H., & Gibbon, J. (1994). Application of scalar timing theory to individual trials. *Journal of Experimental Psychology: Animal Behavior Processes, 20*, 135-155. doi: 10.1037//0097-7403.20.2.135

Fraisse, P. (1963). *The Psychology of Time*. New York: Harper & Row.

Gamache, P.-L., & Grondin, S. (2010). Sensory-specific clock components and memory mechanisms: Investigation with parallel timing: Time, memory and multisensory processing. *European Journal of Neuroscience, 31,* 1908-1914. doi: 10.1111/j.1460-9568.2010.07197x

Gibbon, J. (1977). Scalar Expectancy Theory and Weber's Law in Animal Timing Psychological Review,84, 279-325. doi: 10.1037/0033-295X.84.3.279

Grondin, S. (1993). Duration discrimination of empty and filled intervals marked by auditory and visual signals. *Perception and Psychophysics, 54*, 383-394. doi:10.3758/BF03205274

Grondin, S. (2010). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception, & Psychophysics, 72,* 561-582. doi: 10.3758/APP.72.3.561

Grondin, S. (2020). *The Perception of Time – Your Questions Answered.* New York: Routledge.

Jones, A. (2014). Independent effects of bottom-up temporal expectancy and top-down spatial attention: An audiovisual study using rhythmic cueing. *Frontiers in Integrative Neuroscience, 8,* doi:10.3389/fnint.2014.00096

Jones, M. R. (1976). Time, our lost dimension: Toward a New Theory of Perception, Attention, and Memory. *Psychological Review, 85,* 323-355. doi:10.1037/0033-295X.83.5.323

Jones, M. R. (2018). Time Will Tell: A Theory of Dynamic Attending. New York, NY: Oxford University Press.

Jones, M. R., & Boltz, M. (1989). Dynamic Attending and responses to time. *Psychological Review, 96,* 459-491. doi: 10.1037/0033-295X.96.3.459

Jones, M. R., & McAuley, J. D. (2005). Time judgments in global temporal contexts. *Perception & Psychophysics, 67,* 398-417. doi:10.3758/BF03193320

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time- varying events. *Psychological Review, 106,* 119-159. doi: 10.1037/0033- 295x.106.1.119

McAuley, J. D. (1995) Perception of time as phase: toward an adaptative oscillator model of rhythmic pattern processing. Indiana University, Bloomington, United-States.

McAuley, J. D., & Jones, M. R. (2003). Modeling effects of rhythmic context on perceived duration: A comparison of interval and entrainment approaches to short-interval timing. *Journal of Experimental Psychology: Human perception and Performance, 29,* 1102-1125. doi: 10.1037/0096-1523.29.6.1102

Nobre, A. C., & van Ede, F. (2017). Anticipated moments: temporal structure in attention. *Nature Reviews Neuroscience*, 19, 34-48. doi: 10.1038/nrn.2017.141

Rammsayer, T. H., Borter, N., & Troche, S. J. (2015). Visual-auditory differences in duration discrimination of intervals in the subsecond and second range. *Frontiers in Psychology, 6*, 7. doi: 10.3389/fpsyg.2015.01626

Rousseau, L., & Rousseau, R. (1996). Stop—reaction time and the internal clock. *Perception & Psychophysics, 58,* 434-448. doi: 10.3758/BF03206819

Psychology Teaching by Using Volumetric Video-Based AR and Avatar AR Technology is More Attractive than those by Face-to-Face and Zoom?[1]

Xuanru Guo[1*], Takashi Yoshinaga[2,3], Aaron Hilton[3], Shuji Harumoto[3], Ema Hilton[3], Shizuka Ono[3], Takeharu Seno[1]
[1]Faculty of Design, Kyushu University, Minami-ku, Fukuoka, 815-8540, Japan
[2]Institute of Systems, Information Technologies and Nanotechnologies, Fukuoka, 815-8540, Japan
[3]Steampunk Digital, Co. Ltd., Fukuoka, 814-0001, Japan
E-mail: [1]guoxuanru1101@gmail.com & seno@design.kyushu-u.ac.jp

Abstract— This is a brief introduction of the submitted paper in PRESENCE: Virtual and Augmented Reality in 2021, entitled Psychology teaching by using Volumetric video augmented reality and Avatar based augmented reality technology is more presence and attractive than those by Face to face and Zoom, authored by Guo, Yoshinaga, Hilton A., Harumoto, Hilton, E., Ono, Seno. Based on previous literature, we developed two new methods; Volumetric video AR technology (VAR), Avatar based AR (AAR). Few studies have compared face to face, online, and AR teaching methods and put into the same research simultaneously. Thus, this study compared four different teaching methods VAR, AAR, Face to face and Zoom, and employed a simple psychology course as the teaching content. We found that there was no significant difference in teaching content effect under four different teaching methods, but VAR and AAR created significantly more presence, novelty, and satisfaction, and teacher's attractiveness than Face to face and Zoom groups.

Keywords— Volumetric video AR technology; Avatar based AR technology; Face to face; Zoom online; Psychology course; Attractive

I. Introduction

This is a brief introduction of the submitted paper in PRESENCE: Virtual and Augmented Reality, entitled Psychology teaching by using Volumetric video augmented reality and Avatar based augmented reality technology is more presence and attractive than those by Face to face and Zoom (Guo, Yoshinaga, Hilton A., Harumoto, Hilton, E., Ono, Seno, 2021). With the development of 3D virtual technology, augmented reality technology (AR) and other new methods are gradually used in the teaching process to improve the teaching effect. AR technology was first developed by Sutherland (1968) using head-mounted devices, and currently it is mainly composed of three parts: display system; 3D registration system and human-computer interaction system (Berryman, 2012; Chen et al., 2019). AR is the process of combining some virtual data, such as information, and real scenes with three-dimensional interaction (Carmigniani et al., 2011; O'Shea, 2011).

Various types of AR (such as helmet display, computer application, etc.) have been successfully applied to highly operable and spatial teaching courses in medicine, mathematics, chemistry, and so on (Abad-Segura et al., 2020; Yilmaz, 2018; Mesia et al., 2016). The advantages of these AR technologies are more obvious than those of traditional online and face-to-face course (Wang, 2017). And there are still some limitations in the application of these AR technologies to practical teaching (Sırakaya & Alsancak, 2020). And, it's unclear how volume video and virtual avatar technology will work in the teaching process. Thus, we developed two new present methods, namely: Volumetric video Augmented Reality (VAR) and Avatar based Augmented Reality (AAR).

In summary, it is found that most of them compare the teaching or learning effects of the two of methods, and few of the four different teaching methods are put into the same research simultaneously. This study compared four different teaching methods (VAR, AAR, Face to face, Zoom online) and employed a psychology courses as the teaching content. We also surveyed participants' evaluation of teachers' attractiveness and course novelty and satisfaction, so as to comprehensively and systematically understand the teaching differences under different methods.

II. General Methods

Participants and Ethics statement— This study compared the differences of four different teaching methods, so four between groups of experimental subjects were used. These four groups were the VAR group (10 subjects, M= 37.8 SD=5.34), the AAR group (10 subjects, M= 38.2 SD=7.29), the Face to face group (12 subjects, M= 34.33 SD=10.95), the Zoom group (12 subjects, M= 41 SD=8.28). All participants had a normal vision and no reported any damage. Experimental procedure was approved in advance by the Ethics Committee of Kyushu University. Participants were informed about the experimental stimuli and procedures prior to the experiment, and informed consents were collected from all participants.

Apparatus and Stimulus— Figure 1 shows the outline of the system for condition VAR and AAR. This system enables remote communication in an environment where the appearance and behavior of teachers and subjects in remote areas are displayed three-dimensionally in front of them.

The VAR system used Azure Kinect (Microsoft, USA, 2019) to capture the user's appearance in real time. Information taken with Azure Kinect on the End Point ① side is compressed into an RGBD image that combines depth and color information in order to reduce the data size. WebRTC technology (Google, USA, 2010) is used for the remote video communications link and for distribution of the combined RGBD video. On the End Point ② side, the received RGBD image was converted back into 3D information in the colored Point Cloud format and displayed in front of the user through the Nreal Light of the AR glass (Nreal, 2019, China). The AAR system used the TrueDepth camera in the iPhone 12 Pro (Apple, USA, 2020) to measure the movement of the face at End Point ①, and the system was developed using the Face Tracking function provided by ARKit (iPhone 12 Pro, Apple, 2020, USA). WebRTC was used as the communication technology to distribute the coordinates of each feature point on the face to remote locations. In End Point ②, the facial expression of the other party was reproduced by reflecting the received facial feature point coordinates on the avatar's face, and AR was displayed in front of the eyes using Nreal Light. In addition, in order to prepare the experimental content for both two systems, the lecture content is recorded in advance as a

[1] This conference paper is based on the full journal paper under reviewed in Presence: Virtual and Augmented Reality

time-series data of volumetric video information (VAR system) and facial movements (AAR system) synced with the audio recording.

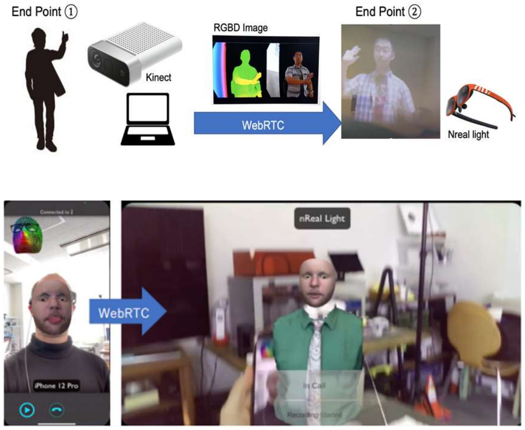


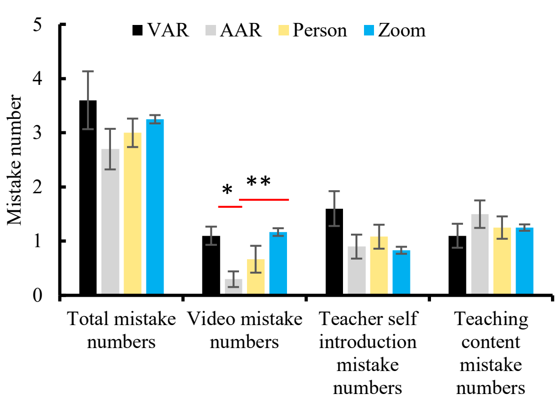
Figure 1. VAR (up) and AAR (down) equipment.

Procedure— After simple communication at the start of the experiment under each condition, the educational contents were watched or taught. After taking the 5 minutes course content,[2] and conducted a questionnaire. The questionnaire prepared 30 questions.[3] It took about 20 minutes for each participant to complete the experiment. The higher the score, the more fun or novelty the course is. That's the way we all score forward, not the other way around. For better statistical analysis, we divide these 14 questions into five types; (1) Course novelty (4 questions: Q15, Q16, Q17, Q19); (2) 3D: (2 questions: Q20, Q21), (3) Video quality (3 questions: Q18, Q23, Q24), (4) Teachers' attractive (2 questions: Q25, Q26), (5) Familiarity (2 questions: Q27, Q28)).

Data analysis—Dependent variables were averaged prior to being subjected to repeated measure ANOVAs. We used the Bonferroni-corrections when conducting multiple post hoc comparisons. All statistical levels were set to 0.05. Partial eta-squared (η_p^2) measures of effect size were also calculated for the ANOVAs. Pearson correlations were used to examine the relationships between parameters.

III. Results

We calculated the responses of participants to different questions in the questionnaire and compared the correct answers to get the number of errors per participant. Figure 2 indicates the number of incorrect questions recalled by participants under different conditions.



[2] https://www.youtube.com/watch?v=_b-rLUe5a4E
[3] https://docs.google.com/forms/d/e/1FAIpQLSdIEKtwqtdW_ZVRw23PQNglFuh4yT3qnP5Ha6WrOo5U5orw/viewf orm?vc=0&c=0&w=1&flr=0&gxids=7628.

Figure 2. Number of falsely recalled questions in teaching or teacher's relevant content of four different teaching methods. * stands p < 0.05, **stands p < 0.01.

From Figure 2, we can see that there is no significant difference in the recall accuracy of the four teaching methods with respect to the total mistake numbers. However, there are significant differences in some methods, such as video mistake numbers.

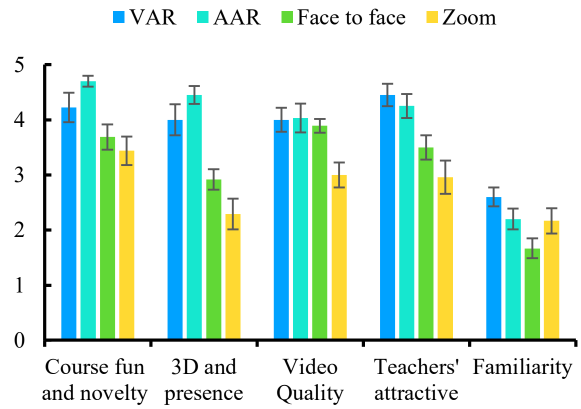


Figure 3. Five aspects of questionnaire result under four different teaching methods.

Figure 3 shows that: in the questionnaire conducted, each item was evaluated on a scale of 1 to 5. In general, we can see that the attractiveness of teacher, 3D and so on are higher under the new technology (VAR, AAR), as follows.

IV. General Discussions

The main purpose of this study is to investigate the differences in teaching content effectiveness and subjective presence and attractiveness under the four different teaching methods. The results showed that there was no significant difference in teaching content effect under the different teaching methods, but there were significant differences in teaching novelty, satisfaction, presence and teacher's attractiveness under these four different teaching methods. Specially, in the VAR and AAR group, participants felt the course more novel and interesting, they also perceived the teacher was more 3D and attractive.

Previously, many studies have also applied AR technologies to the teaching process and found that with these AR technologies, the learning effect has been improved (Cai et al, 2014). These previous results are partially consistent with the results of this study, which also found that there was no difference in participants' recall of the teaching content when using the new VAR and AAR technology. But effected more presence and attractiveness in teaching of VAR and AAR groups. This is consistent with previous research results, once again prove that the new AR technology will improve student satisfaction and participation (Huang et al, 2016; Mesia et al., 2016). In other words, the new VAR and AAR technology developed can also improve students' interest in learning and increase course satisfaction. In summary, we can find that the new VAR and AAR technologies that we have developed can enhance course satisfaction, teacher's attractiveness and improve the presence and engagement of simple fun psychology courses.

References

Abad-Segura, E., González-Zamar, M. D., Rosa, A. L. D. L., & Cevallos, M. B. M. (2020). Sustainability of educational technologies: An approach to augmented reality research. *Sustainability*, 12(10), 4091

Berryman, D. R. (2012). Augmented reality: a review. *Medical Reference Services Quarterly*, 31(2), 212-218.

Carmigniani, J., Furht, B., Anisetti, M., Ceravolo, P., Damiani, E., & Ivkovic, M. (2011). Augmented reality technologies, systems and applications. *Multimedia Tools and Applications*, 51(1), 341-377.

Cai, S., Wang, X., & Chiang, F. K. (2014). A case study of Augmented Reality simulation system application in a chemistry course. *Computers in Human Behavior*, 37, 31-40.

Chen, Y., Wang, Q., Chen, H., Song, X., Tang, H., & Tian, M. (2019, June). An overview of augmented reality technology. In *Journal of Physics: Conference Series* (Vol. 1237, No. 2, p. 022082). IOP Publishing.

Guo, X.R., Yoshinaga, T., Hilton, A., Harumoto, S., Hilton, E., Ono, S., Seno, T. (2021). Psychology teaching by using Volumetric video augmented reality and Avatar based augmented reality technology is more presence and attractive than those by Face to face and Zoom. PRESENCE: Virtual and Augmented Reality. (under review)

Huang, H. M., Liaw, S. S., & Lai, C. M. (2016). Exploring learner acceptance of the use of virtual reality in medical education: a case study of desktop and projection-based display systems. *Interactive Learning Environments*, 24(1), 3-19.

Mesia, N. S., Sanz, C., & Gorga, G. (2016). Augmented reality for programming teaching. Student satisfaction analysis. In 2016 international conference on collaboration technologies and systems (cts) (pp. 165-171). IEEE.

O'Shea, P. M. (2011). Augmented reality in education: current trends. International *Journal of Gaming and Computer-Mediated Simulations*, 3(1).

Sırakaya, M., & Alsancak Sırakaya, D. (2020). Augmented reality in STEM education: A systematic review. *Interactive Learning Environments*, 1-14.

Wang, Y. H. (2017). Using augmented reality to support a software editing course for college students. *Journal of Computer Assisted Learning,* 33(5), 532-546.

Yilmaz, R. M. (2018). Augmented reality trends in education between 2016 and 2017 years. State of the art virtual reality and augmented reality knowhow, 81, 97.

Vection Depends on the Luminance Contrast, Average Luminance and the Spatial Frequency of the Stimulus

Xuanru Guo[1*], Shinji Nakamura[2], Yoshitaka Fujii[3], Takeharu Seno[1*], Stephen Palmisano[4]
[1]Faculty of Design, Kyushu University, Minami-ku, Fukuoka, 815-8540, Japan
[2]School of Psychology, Nihon Fukushi University, Okuda, Mihama-cho, Aichi 470-3295, Japan
[3]Graduate School of Humanities and Social Sciences, Kumamoto University, Kumamoto, 860-8555, Japan
[4]School of Psychology, University of Wollongong, Wollongong, NSW 2522, Australia
E-mail: [1]guoxuanru1101@gmail.com & seno@design.kyushu-u.ac.jp

Abstract— This is a brief introduction of the accepted paper in Experimental Brain Research in 2021, entitled Effects of luminance contrast, averaged luminance and spatial frequency on vection, authored by Guo, Nakamura, Fujii, Seno, Palmisano. The size, speed, and even the perceived material properties, of global visual motion stimuli are all known to affect the intensity of vection (i.e., visually induced illusions of self-motion). To date, there has been not been a systematic study of the effects of luminance contrast, averaged luminance and stimulus spatial frequency on vection. This study therefore examined the vection induced by downward drifting grating stimuli with six different levels of luminance contrast (from 0.046 to 0.999), four different levels of averaged luminance (from 1.59 to 17.035 cd/m^2) and four different spatial frequencies (from 0.02 to 5 c/deg). The first three experiments demonstrated that vection intensity is altered by manipulating each of these visual stimulus properties – specifically they showed that vection increased with either the grating's luminance contrast, its averaged luminance, or its spatial frequency. The results from a fourth experiment showed that there are also complex interactions between the effects of these three factors. While motion energy modelling suggested that these behavioural vection data could arise during low level visual processing, higher level effects on perceptions (e.g., of stimulus visibility, brightness or stimulus speed) could also have contributed these effects.

Keywords— Vection; Luminance Contrast; Averaged Luminance; Spatial Frequency; Speed; Motion Perception

I. Introduction

This is a brief introduction of the accepted paper in Experimental Brain Research, entitled Effects of luminance contrast, averaged luminance and spatial frequency on vection (Guo, Nakamura, Fujii, Seno, Palmisano, 2021). Vection refers to the visual illusion of self-movement produced by the observer when observing the surrounding movement in a state of rest (Palmisano et al., 2015). Previous studies have found that vection is influenced by a number of physical properties of visual stimuli, for example frame rate (Fujii et al., 2019), size (Nakamura, 2006), speed (Brandt et al., 1973), positions in the visual filed (Nakamura, & Shimojo., 1998), smoothness of stimulus (Fujii, Seno, & Allison, 2017) and so on.

While the effects of these stimulus properties have been studied in detail, there has been less investigation of the effects of luminance contrast (i.e., the ratio between the maximum and the minimum luminance of the stimulus), absolute luminance level (i.e., the average luminance across the entire visual stimulus) and stimulus spatial frequency on vection. Importantly, there has been even less examination of how these different properties might interact to determine vection. More recent research has tended to use computer generated display motions (Patterson and York 2009; Sauvan and Bonnet 1993, 1995; Gurnsey et al. 1998). In terms of the effects of stimulus spatial frequency, research suggests that vection (like perceived velocity) generally increases with the number/density of moving contrasts (Brandt et al. 1973; Dichgans and Brandt 1978; Sauvan and Bonnet 1993; Palmisano and Gillam 1998).

In order to investigate the unique influence of each of these visual properties on vection, we manipulated levels of either the motion stimulus' luminance contrast (in Experiment 1), its averaged luminance (in Experiment 2), or its spatial frequency (in Experiment 3), while holding the two other factors constant. We then investigated the possible interactions between these three factors in Experiment 4 (using a fully factorial design, where all combinations of the levels for each factor were tested, to determine the relative contribution of each visual property to the vection experience). There is may be able to explain these effects of such stimulus manipulations on vection based simply in terms of motion energy. Alternatively, it is possible that they might be better explained by changes to the perceived qualities of the inducing stimuli.

II. General Methods

Participants and Ethics statement— Eighteen college students (8 females and 10 males) volunteered to participate in first three experiments. Their ages ranged from 21 to 37 years (*M*= 26.67 *SD*=4.18). Eighteen college students (8 females and 10 males), who had not been involved in Experiments 1-3, volunteered to participate in this Experiment 4. Their ages ranged from 21 to 35 years (*M* = 26.33, *SD* =4.15). All had normal or corrected-to-normal vision and no reported history of either attention deficit disorder, mental illness, or brain damage. The experimental procedure was approved in advance by the Ethics Committee of Kyushu University.

Apparatus and Stimulus— The experimental stimulus were presented on Plasma display (3D Viera 65-inch, Panasonic, Japan, with 1920 × 1080 pixel resolution at a 60 Hz refresh rate) and controlled by a computer (MacBook Pro Retina, 13-inch, Mid 2014). The experimental program was programmed by MATLAB R2014a (Mathworks, Natick, MA), which also used PsychToolbox-3 (Brainard, 1997), and the fixpoint and rating slide were presented on a gray background (RGB: 127,127,127). This study used downward moving grating stimuli with for 30 seconds. Participants participated in the test in a dark laboratory, and participants viewed the screen from a distance of 65 cm.

Procedure—The flow chart of the experiment is shown in Figure 1.

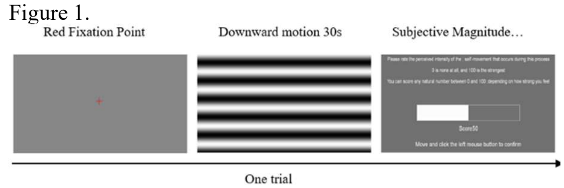


Figure. 1. Schematic representation of the experimental procedures in these experiments.

The flow of a single trial in the experiment is shown in Figure. 1. First, a red fixation point(size:2°) is displayed for 1 second. Then the motion grating stimulation appeared for 30 seconds. During the process of motion stimulation, if the participants felt that they were moving, they had to press the space bar all the time. If they did not feel that they were moving, they did not press it. Finally, a scoring interface appeared on the display so that participants could estimate magnitude of their vection for that trial. They were also asked to rate other perceived attributes of the grating stimulus used on that trial, such as its visibility (Experiments 1 and 4), its perceived intensity/brightness (Experiments 2 and 4) and its perceived speed (Experiments 1-3). All of the scores were from 0 to 100. For vection magnitude ratings, 0 indicated "no self-motion was perceived" and 100 indicated "very strong and saturated self-motion was perceived". For their visibility ratings, 0 indicated "the grating was very hard to perceive" and 100 indicated "the grating was perceived very clearly".

Data analysis—Dependent variables were averaged prior to being subjected to repeated measure ANOVAs. Pearson correlations were used to examine the relationships between parameters (between three vection strength indices and grating visibility, brightness, speed, simulations of motion energy).

III. Results

Experiment 1: Effect of Luminance Contrast— The primary aim of Experiment 1 was to investigate the effect of luminance contrast on the experience of vection. It had the following 2 (Grating Speed) by 6 (Luminance Contrast) within-subjects design – resulting in 12 different stimulus conditions. On each trial, the grating stimulus drifted down at one of two different speeds, either 20°/s or 60°/s. The grating stimulus also had one of six different levels of luminance contrast: 0.046, 0.139, 0.245, 0.469, 0.755, or 0.999. Each of the 12 experimental conditions was repeated four times, and thus each participant underwent 48 experimental trials in a randomized order (4 repetitions of the 12 different stimulus conditions). Testing required approximately 30 minutes. The visual stimuli employed in this experiment and results are shown in Figure 2.

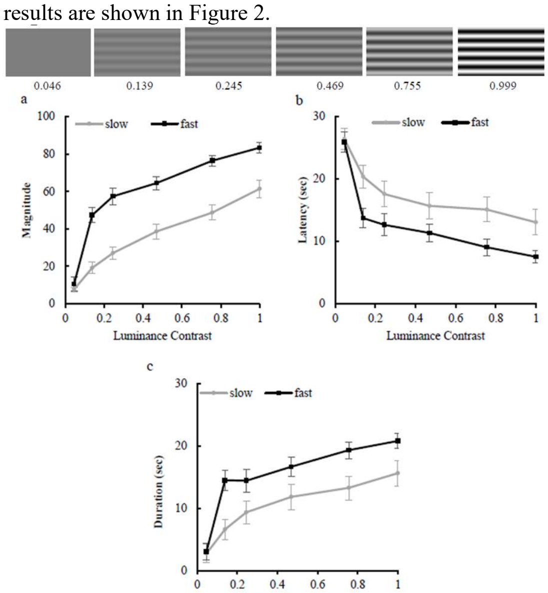


Figure 2: Gratings with six different levels of luminance contrast used in Experiment 1. Average vection magnitude(a); vection latency(b); duration(c) as a function of the grating's luminance contrast and motion.

Figure 2 shows the average vection magnitudes, latencies and durations as a function of the luminance contrast in each of the stimulus-speed conditions. The results clearly show that vection magnitude and duration both increased in a linear fashion with the luminance contrast. The latency to vection onset can also be seen to decrease in a linear fashion as the luminance contrast increased. These results also show that vection was more compelling in the 60°/s (compared to 20°/s) conditions. The grating visibility was reported to increase linearly with the luminance contrast, in a similar fashion to the vection strength measures. By contrast, there were no clear effects of luminance contrast on perceived grating speed.

Experiment 2: Effect of Averaged Luminance— We examined the effects of four different averaged luminance conditions, namely 1.590, 3.060, 9.200, 17.035 cd/m^2, as show in Figure 3. We again examined the effects of two different grating speeds: 20°/s and 60°/s (as in Experiment 1). Thus, there were eight different stimulus conditions: 2 (Grating Speeds) x 4 (Averaged Luminance Levels). Each of the 8 experimental conditions was repeated four times, and thus each participant underwent 32 experimental trials in a randomized order.

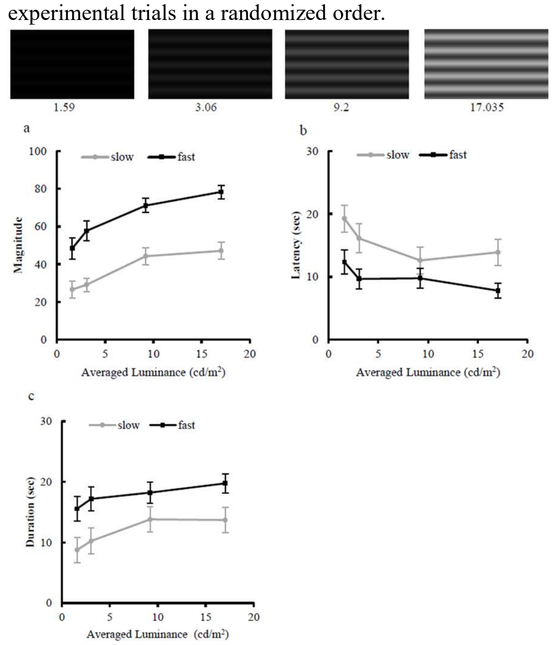


Figure 3: Gratings with four different levels of averaged luminance used in Experiment 2. Average vection magnitude(a); vection latency(b); duration(c) as a function of the grating's averaged luminance and motion.

Figure 3 shows the average vection magnitudes, latencies and durations as a function of the averaged luminance for each of the stimulus speed conditions. The results clearly indicate that vection strength increases linearly with the averaged luminance; estimated magnitudes became higher, durations became longer and latencies became shorter in higher averaged luminance conditions. And the subjective brightness increased linearly with the averaged luminance of the grating, in a similar fashion to the relationships observed with the three vection strength indices. There were no clear effects of the averaged luminance on the perceived speed of the visual stimulus. However, faster grating motions did increase the ratings of both brightness and perceived speed.

Experiment 3: Effects of Stimulus Spatial Frequency— Experiment 3 investigated the effects of stimulus spatial frequency on vection. Four different spatial frequency conditions were examined, namely 0.02, 0.15, 1.25 and 5 c/deg (as shown in Figure 4).

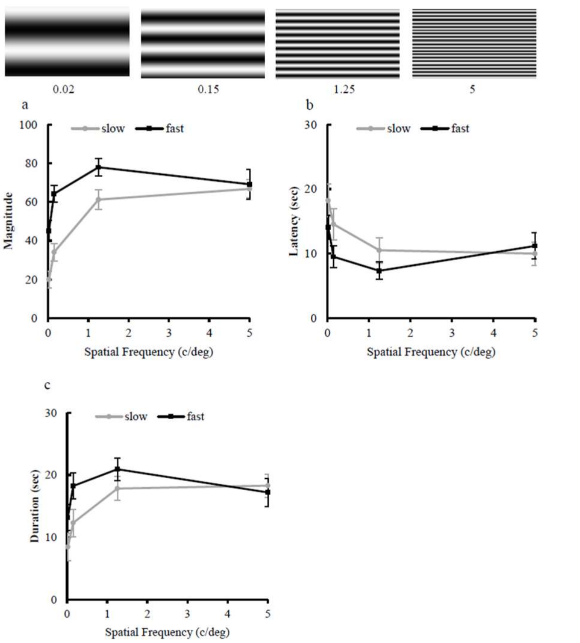


Figure 4: Gratings with four different levels of spatial frequency used in Experiment 3. Average vection magnitude(a); vection latency(b); duration(c) as a function of the grating's spatial frequency and motion.

Figure 4 shows the average vection magnitudes, latencies and durations as a function of stimulus spatial frequency under each of the two speed conditions. Vection strength can be seen to increase in linear fashion with the spatial frequency of the gratings; with vection magnitudes being higher, vection durations being longer, and vection latencies being shorter in higher spatial frequency conditions. As in Experiments 1 and 2, vection was typically stronger during faster grating motion. However, we observed no effect of stimulus speed on vection strength for the highest spatial frequency condition. No clear effects of spatial frequency were found for the perceived speeds of these drifting gratings.

Experiment 4— Two levels (low and high) of each the three different independent variables were examined: 1) 2 levels of averaged luminance: 2.523 and 14.891 cd/m^2; 2) 2 levels of luminance contrast: 0.048 and 0.625, and 3) two levels of spatial frequency: 0.02 and 1.25 c/deg. Unlike Experiments 1-3, only one stimulus speed: 60 °/s. The experiment procedure was otherwise the same as that for Experiment 1.

Experiment 4 shows participants' average vection magnitudes, latencies and durations as a function of stimulus luminance contrast, averaged luminance level and spatial frequency. In general, vection strength was again found to increase in a linear fashion with each of these factors; vection magnitudes were higher, vection durations were longer and vection latencies were shorter in higher luminance contrast, higher averaged luminance and higher spatial frequency conditions. Importantly, interactions between these three factors were found for the vection induced in Experiment 4. Visibility and brightness ratings can be seen to increase with the luminance contrast and the averaged luminance.

IV. General Discussions

The main motivation of this study was to systematically investigate the effects of luminance contrast, averaged luminance and stimulus spatial frequency on vection. These three factors are generally agreed to be fundamental to the processing of all visual stimuli (Cai et al. 2019; Shi et al. 2020). Four psychophysical experiments revealed that all three factors had significant impacts on vection.

From the point of view of the motion energy, we calculated the motion energy simulation values under different conditions. Higher luminance contrast and higher averaged luminance can be seen to elicit stronger motion energy. The results of these simulations appear quite similar to the behavioural results of our experiments, where vection strength was also found to be stronger in higher luminance contrast (Experiment 1) and higher averaged luminance (Experiment 2) conditions. Thus, it is possible that these vection strength effects directly simply reflect the motion energy of the low-level motion detectors.

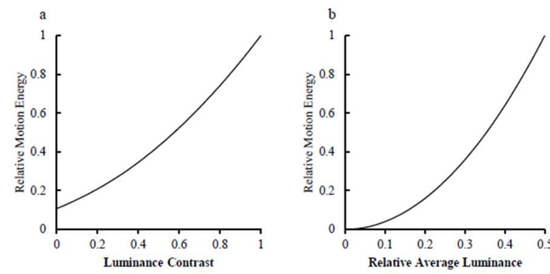


Figure 5: Results of simulations based on the motion energy model. (a): luminance contrast. Averaged luminance was constant (= 0.5). (b): Results: averaged luminance. Luminance contrast was constant (= 1).

Our manipulations of luminance contrast, averaged luminance, spatial frequency and motion speed, did not only alter the vection strength, but also the perceived qualities of the visual inducer, such as its visibility, its perceived intensity/brightness and its perceived speed. Participant ratings of the visibility and the brightness of these drifting grating stimuli were found to increase (in a linear fashion) with their luminance contrast and their averaged luminance, respectively. In summary, these three factors each had significant effects on vection – which could potentially be explained by their effects on motion energy of the inducing stimuli.

References

Brainard, D. H. (1997). The psychophysics toolbox. Spatial Vision, 10, 433–436.

Brandt, T., Dichgans, J. & Koenig, E. (1973) Differential effects of central versus peripheral vision on egocentric and exocentric motion perception. Experimental Brain Research, 5, 476-491.

Cai, L. T., Yuan, A. E., Backus, B.T. (2019) Binocular global motion perception is improved by dichoptic segregation when stimuli have high contrast and high speed. Journal of Vision 19(13):10.

Dichgans, J, Brandt, T. (1978) Visual-Vestibular Interaction: Effects on Self-Motion Perception and Postural Control. In Perception, Handbook of Sensory Physiology (pp. 756 - 804). Springer.

Fujii, Y., Kio, O. G., Au, D., Wilcox, L. M., & Allison, R. S. (2019). Effects of Frame Rate on Vection and Postural Sway. Displays.

Fujii, Y., Seno, T., & Allison, R. S. (2017). Smoothness of stimulus motion can affect vection strength. Experimental Brain Research, 236(1), 243–252.

Guo, X.R., Nakamura, S., Fujii, Y., Seno, T., Palmisano, S. (2021). Effects of luminance contrast, averaged luminance and spatial frequency on vection. Experimental Brain Research. (In press) doi:10.1007/s00221-021-06214-5

Gurnsey, R., Fleet, D., & Potechin, C. (1998). Second-order motions contribute to vection. Vision Research, 38(18), 2801–2816.

Nakamura, S., & Shimojo, S. (1998). Stimulus Size and Eccentricity in Visually Induced Perception of Horizontally Translational Self-Motion. Perceptual and Motor Skills, 87(2), 659–663.

Nakamura, S. (2006). Effects of depth, eccentricity and size of additional static stimulus on visually induced self-motion perception. Vision Research, 46(15), 2344–2353.

Palmisano, S., Allison, R. S., Schira, M. M., & Barry, R. J. (2015). Future challenges for vection research: definitions, functional significance, measures, and neural bases. Frontiers in psychology, 6, 193.

Palmisano, S, Gillam, B. (1998) Stimulus eccentricity and spatial frequency interact to determine circular vection. Perception 27(9):1067-1077.

Patterson, F., and York, Y. (2009). Vection and Motion Thresholds as a Function of Contrast. Technical Report No. NAMRL-09-32. Pensacola, FL, USA: Naval Aerospace Medical Research Lab.

Sauvan, X. M., & Bonnet, C. (1993). Properties of curvilinear vection. Perception & Psychophysics, 53(4), 429.

Sauvan, X. M., & Bonnet, C. (1995). Spatiotemporal boundaries of linear vection. Perception & Psychophysics. 1995,57 (6). 898-904.

Seno, T., Palmisano, S., Ito, H., & Sunaga, S. (2012). Vection can be induced without global-motion awareness. Perception,41(4), 493-497.

Shi, C., Pundlik, S., Luo, G. (2020) Without low spatial frequencies, high resolution vision would be detrimental to motion perception. Journal of Vision 20(8):29.

How do We Perceive Masked Faces during COVID 19 Pandemic?

Evrim Gülbetekin,[1] Arda Fidancı,[2] Enes Altun,[1] Muhammed Nurullah Er,[1] & Elif Ezgi Kaplan[1]
[1]Department of Psychology, Akdeniz University
[2]Cognitive Science Department, University of Minnesota
evrimg@akdeniz.edu.tr

Abstract— The face provides a rich source of information about its owner. Following the introduction of public health measures to respond to COVID-19, wearing a mask has become one of the most important precautions. However, considerable visual information has become limited due to masks coverage of the majority of the face. We investigated both how mask use affected face recognition and the recognition of facial expressions during the COVID 19 pandemic. We used a matching-to-sample procedure in the first experiment. Participants exhibited the best performance in recognizing an unmasked face condition and the poorest when asked to recognize a masked face that they had seen earlier without a mask. In the second experiment we investigated effect of mask on emotion recognition. We observed that emotion recognition performance decreased for faces portrayed with masks. The emotional expressions ranged from the most to least accurately recognized as follows: happy, neutral, disgusted, and fearful. In Experiment 3 we aimed to compare perception of masked faces and the faces having threat-related expressions. We presented two facial stimuli simultaneously: one face portrayed with a mask and one with an emotional expression (fear, disgust or neutral) and asked participants to indicate which face they perceived at first. We used electroencephalographic analysis of event-related potentials (ERPs) to investigate how the brain compares masked faces with emotional faces. The preliminary analysis of P300 component indicated that the task activated occipital and parietal regions however, we did not observe any difference among emotional conditions.

Keywords— face perception, recognition of facial expressions, mask use, electrophysiology

I. INTRODUCTION

The face provides a rich source of information about its owner. However, following the spread of the COVID-19 wearing a procedure mask has become obligatory in many places. Given that masks cover large areas of the face, including the nose, mouth, chin, and most of the cheeks, visual information has become limited in face perception. Prior to the current pandemic, researchers indicated impairing effects of facial parts occlusion in the context of face recognition and recognition of emotional expressions (Bassili 1979; Roberson et al., 2012; Dhamecha et al. 2014). More recently, researchers showed that face recognition performance (Carbon, 2020) (Freud et al. 2020), face matching performance and emotion recognition performance (Carragher & Hancock, 2020) decreased when participants were exposed to faces with masks during the COVID-19 pandemic.

Until the COVID-19 pandemic, only people who had specific medical reasons were likely to wear face-covering masks. Therefore, at the beginning of the COVID-19 pandemic, wearing facemasks might have led to a negative perception of mask wearers. However, mask use became more prevalent ongoing pandemic period since health authorities informed the public that facial masks increased the protection against COVID-19. Therefore, it is thought that perception toward people wearing facial masks might change during the pandemic. Facial masks might pose a threat or a safety signal just like emotional expressions to adjust social distance with other people given the likely association between a masked face and fears of contamination. There may be other potential hypotheses: It may create a noise for signal detection, or it may be associated with negative affect due to low frequency of it's use.

To examine the effect of mask use, the current study consisted of three experiments to test the effect of mask use on face perception in three contexts during the COVID-19 pandemic: (a) how mask use influences face recognition, (b) how mask use influences the recognition of facial expressions, and (c) perception of masked faces compared to facial expressions.

EXPERIMENT I

Method

Participants

A total of 102 undergraduate students (80 females and 22 males) with a mean age of 20.4 years (M = 20.4, SD = 2.8) participated in the study.

Stimuli

Twenty-four Asian and 24 Caucasian facial stimuli (24 female and 24 male) were used from the racially diverse affective expression (RADIATE) face stimulus set (Conley et al., 2018) and the MR2 (a multi-racial mega-resolution database of facial stimuli; Strohminger, 2016). We used neutral faces in Experiment 1.

Surgical face mask images were superimposed on all of the facial stimuli using Adobe Photoshop software to create faces with masks. The same surgical face mask image was used for all the masked stimuli. The mask was positioned to cover each image's chin, nose, and most of the cheek area.

Procedure

Experiment 1 involved a matching-to-sample procedure. A face without a mask was presented, followed by two faces (including the target face). The participants were asked to choose the face that had been presented as a sample. The experimental sessions incorporated four conditions: 1) The sample was an unmasked face; the test faces were also unmasked. 2) The sample face was masked; the test faces were also masked. 3) The sample face was unmasked, but the test faces were masked. 4) The sample face was masked, but the test faces were unmasked.

After a fixation cross was presented for 500 ms, the sample face was presented for 500 ms. Following sample face presentation, two faces were presented for 800 ms, and the participants were asked to decide which face was the sample. The participants pressed "1" on the keyboard to choose the stimulus on the left side of the screen and "0" to choose the stimulus on the right side of the screen.

Due to the pandemic conditions, data collection employed an online platform. Fifty participants were tested using the online experiment platform Testable, and 52 participants were tested via E-Prime Go.

Results

Accuracy

A 4 × 2 × 2 (mask conditions [(1) unmasked sample–unmasked test, (2) masked sample– masked test, (3) unmasked sample–masked test, (4) masked sample–unmasked test] × race [Asian, Caucasian] × sex [female, male]) mixed ANOVA was conducted to determine the effects of mask conditions, race, and sex of facial stimuli on discrimination accuracy. The results were presented in Table 1 and Table 2.

Table 1
Mixed ANOVA Results for Accuracy

Predictor	df_{Num}	df_{Den}	*Epsilon*	SS_{Num}	SS_{Den}	*F*	*P*	ηp^2
Intercept	1	101		1440.38	46661	141.25	.000	.99
Race	1	101	1	.94	44317	90.28	.000	.47
Mask condition	44349	263.08	.87	.46	.008	60.79	.000	.38
Race × Mask condition	22313	263.50	.87	.345	.009	37.84	.000	.27
Sex × Mask condition	27061	227.49	.92	.261	.009	29.59	.000	.23
Race × Sex × Mask condition	3	303	.95	.131	.008	16.33	.000	.14

Table 2

Descriptive Statistics (Means and Standard Deviations) for Accuracy in Respect to Mask Conditions

	M	*SD*
Unmasked sample–Unmasked test	.97	.09
Unmasked sample–Masked test	.89	.07
Masked sample–Unmasked test	.94	.09
Masked sample–Masked test	.95	.09

The results indicated a significant mask impact on accuracy. The participants' performance was the best when they were asked to recognize an unmasked face that they had seen previously. However, their performance decreased when they first saw an unmasked face and then had to recognize that face with a mask. The participants exhibited the best performance for unmasked faces and the poorest when asked to recognize a masked face that was seen earlier without mask. Race was another critical factor in face recognition. The participants recognized Caucasian faces better than Asian faces qualified by a significant race and mask interaction. Performance was lowest when Caucasian sample mask faces were presented before unmasked test faces. In comparison, performance was lowest when Asian sample faces were presented with a sample unmasked face and a masked test face.

EXPERIMENT II

Method

Participants

A total of 134 undergraduate students (105 females and 29 males) with a mean age of 21 years (M = 21.3, SD = 1.6) participated in the study. All participants were tested using the online experiment platform Testable.

Facial stimuli were chosen from the racially diverse affective expression (RADIATE) face stimulus set. In total, we used eight Asian (four female, four male) and eight Caucasian (four female, four male) faces. Four emotional expressions (neutral, happy, fear, and disgust) were used among the 16 faces. Since our aim was to test the effect of mask use on emotion recognition, surgical face mask images were superimposed on all facial stimuli using Adobe Photoshop software, as in Experiment 1. The same surgical face mask image was used for all the stimuli. In total, we had 128 experimental stimuli: 64 unmasked faces and 64 masked versions of those faces.

A fixation cross was presented for 500 ms. Next, a neutral face or a face with an emotional expression (happy, fear, disgust) was presented on the center of the screen for 1,000 ms. The participants were asked to indicate which expression appeared on the presented face. Seven options (Happy, Sad, Surprised, Neutral, Fearful, Disgusted, Angry) were presented to choose. The experiment consisted of 128 trials in total. The stimuli were presented in pseudo-random order.

Results

Accuracy

A 2 × 2 × 2 × 4 (mask conditions [unmasked, masked] × race [Asian, Caucasian] × sex [female, male] × emotion [neutral, happy, fear, disgust]) mixed ANOVA was conducted to determine the effects of mask wearing, emotions, race, and sex of facial stimuli on discrimination accuracy. Since the sphericity assumption was violated, the Greenhouse–Geisser correction was applied in the required analysis. Descriptive statistics are presented in Table 3 and within-subject factor results are presented in Table 4. The results indicated a significant main effect of mask, race, emotion, and stimulus sex.

Table 3
Mixed ANOVA Results for Accuracy

	Neutral unmasked	Neutral masked	Happy unmasked	Happy masked	Fear unmasked	Fear masked	Disgust unmasked	Di ma
Asian female	.64 (.52)	.66 (.24)	.94 (.17)	.66 (.23)	.44 (.22)	.71 (.22)	.74 (.19)	.28
Asian male	.80 (.23)	.76 (.24)	.90 (.20)	.60 (.24)	.36 (.27)	.25 (.24)	.69 (.19)	.16
Caucasian female	.75 (.24)	.59 (.18)	.96 (.17)	.84 (.25)	.52 (.26)	.41 (.26)	.83 (.18)	.39
Caucasian male	.79 (.20)	.86 (.19)	.94 (.19)	.78 (.23)	.64 (.21)	.66 (.28)	.97 (.12)	.44

Table 4. Mixed ANOVA Results for Accuracy

Predictor	df_{Num}	df_{Den}	*Epsilon*	SS_{Num}	SS_{Den}	*F*	*p*
Intercept	1	132		1770.19	24.78	9428.55	.000
Sex	1	132	1	1.25	5.67	28.95	.000
Race	1	132	1	18.65	6.23	395.35	.000
Emotion	2.74	361.1	.91	93.18	59.98	205.06	.000
Mask	1	132	1	43.38	7.25	789.96	.000
Race × Sex	1	132	1	3.83	6.008	84.18	.000
Race × Emotion	2.51	331.4	.90	5.59	16.49	44.75	.000
Sex × Emotion	2.51	331.4	.84	5.61	17.40	42.52	.000
Race × Sex × Emotion	2.51	331.2	.84	2.55	15.90	21.19	.000
Sex × Mask	1	132	1	.50	3.79	17.45	.000
Race × Sex × Mask	1	132	1	.88	4.02	28.89	.000
Emotion × Mask	2.94	388.7	.98	33.88	21.24	210.52	.000
Race × Emotion × Mask	2.67	352.9	.89	1.42	15.58	12.07	.000
Sex × Emotion × Mask	2.66	351.4	.89	1.08	14.03	10.16	.000
Race × Emotion × Mask × Sex	2.85	376.6	.95	.93	12.18	10.02	.000

From the most accurate to the least accurate, the participants recognized the emotional expressions under consideration as follows: happy, neutral, disgusted, and fearful. Mask wearing decreased recognition accuracy for all emotional expressions. Although the order of recognition accuracy for the various emotions did not change for masked faces, the accuracy for happy and neutral masked faces did not differ significantly.

Race was also a crucial factor in emotion recognition. The participants could recognize all Caucasian expressions better than the corresponding Asian expressions. However, the same recognition order was valid for both Caucasian and Asian expressions: happy, neutral, disgusted, and fearful (from the most to the least accurately recognized). Also, the participants recognized male Caucasian expressions better than female Caucasian expressions. However, the recognition pattern was the opposite for Asian faces.

EXPERIMENT III

Method

Participants

A total of 20 volunteer participants aged between 18-35 (8 female and 16 male) were included in Experiment 3. All participants were selected from individuals with normal or corrected normal visual acuity. Participants are people who have not been diagnosed with a neurological-psychiatric or hormonal disorder and do not use any psychiatric drugs that can affect their cognitive processes.

Stimuli

Seven male and 7 female faces were selected from Boğaziçi face database consisting various facial expressions (angry, disgusted, scared, neutral, happy) of Turkish sample. Four different facial expressions (neutral, aggressive, fearful, and disgusted) of each face were used as experimental stimuli. We also created an extra facial stimulus by superimposing a surgical face mask image on the neutral facial expressions by using Adobe Photoshop. The mask covered chin, nose, and most of the cheek areas of the faces.

Procedure

E-Prime 3 software was used for stimulus presentation. A fixation cross was presented for 1000 ms. Then two facial stimuli were presented on the right and the left sides of the screen for 3000ms. The participants were asked to indicate which of the two faces they perceived at first sight by using keyboard ("1" for the face on the left and "0" for the face on the right).

All combinations of five categories (neutral, aggressive, fearful, disgusted, and masked) were presented. Participants were exposed to 200 experimental trials.

Results

ERP

We analyzed the ERP in trials which a masked and unmasked face was matched. The averaged epoch was 1000 ms, including a 200 ms pre-stimulus baseline. P300 components were measured and analyzed. 8 electrode sites (O1, Oz, O2, Pz, P3, P4, CP1 and CP2) were selected for the statistical analysis of P300 component (250–500 ms). A two-way repeated measure analysis of variance (ANOVA) on the amplitude of P300 component was conducted with Face conditions (four levels: angry, fearful, disgusted and neutral), and Electrode site as within-subject factors. We found a significant main effect of Electrode [$F(7, 462) = 7.68$, $p < 0.001$]. The effect of condition was not significant. Specifically, O1, O2, and P4 electrodes elicited larger amplitudes than CP1 and CP2 ($p < 0.05$).

The grand average and scalp topography for four conditions are different (Figure 1). Therefore, we plan to conduct further analyses on N1, P1, VPP, N170, and N300 components to understand the difference among those conditions.

DISCUSSION

Face perception and social distancing have significantly changed over the course of the COVID-19 pandemic. Seeing and recognizing masked faces as well as trying to understand emotional expressions have become common challenges of daily life. Mask wearing decreased both face recognition and recognition of emotional expressions. From the preliminary ERP findings and topographic maps for different conditions, we may infer that even if masked faces have become associated with pandemic, they are not perceived as threat signals such as fearful, angry, or disgusted faces. The task activated attentional network in the brain including occipital and parietal regions in all stimulus conditions. However, scalp topography changed when masked faces were compared with angry, disgusted, or fearful faces. Although occlusion of some facial parts decreased face recognition and emotion recognition performance, it seems that masked faces can still hold the attention.

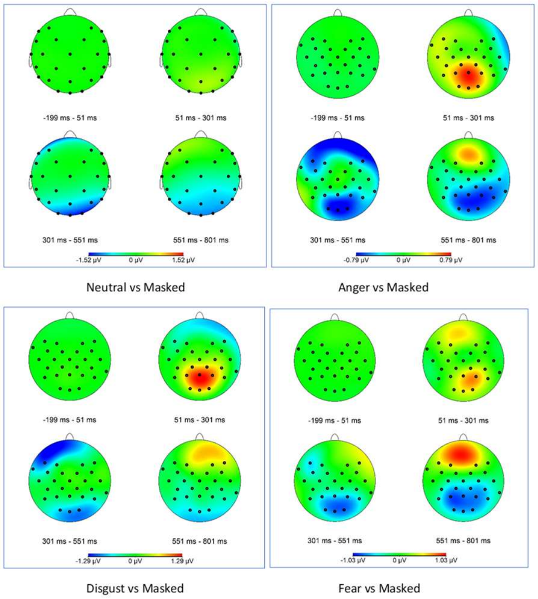


Figure 1. The grand average and scalp topography for four conditions

References

Bassili, J. N. (1979). Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology, 37(11)*, 2049.

Carbon, C.-C. (2020). Wearing face masks strongly confuses counterparts in reading emotions. *Frontiers in Psychology, 11*, 2526.

Carragher, D. J., & Hancock, P. J. B. (2020). Surgical face masks impair human face matching performance for familiar and unfamiliar faces. *Cognitive Research, 5*, Article 59.

Dhamecha, T. I., Singh, R., Vatsa, M., & Kumar, A. (2014). Recognizing disguised faces: Human and machine evaluation. *PLoS One, 9(7),* e99212.

Freud, E., Stajduhar, A., Rosenbaum, R. S., Avidan, G., &Ganel, T. (2020). The COVID-19 pandemic masks the way people perceive faces. *Scientific Reports, 10*, 22344.

Strohminger, N., Gray, K., Chituc, V., Heffner, J., Schein, C., & Heagins, T. B. (2016). The MR2: A multi-racial, mega-resolution database of facial stimuli. *Behavior Research Methods, 48(3)*, 1197–1204.

Higher-Order and Medial Psychophysics: A New Typology

Timothy L. Hubbard
Department of Psychology, Arizona State University
College of Doctoral Studies, Grand Canyon University
timothyleehubbard@gmail.com

Abstract — A typology involving higher-orders of psychophysics and a new domain of medial psychophysics are proposed. First-, second-, and third-order psychophysics involve relationships of subjective experiences or brain states to perceived, remembered, and imputed physical properties, respectively, of a stimulus. Fourth-order psychophysics involves relationships of subjective experiences or brain states to agentive properties of a stimulus. Characteristics and examples of each order are provided. Medial psychophysics involves relationships between properties of a stimulus and brain states. It is suggested that such a typology can provide a structure for bringing a range of phenomena previously viewed as unrelated into a common theoretical framework and can extend the range of possible application of psychophysical principles, theories, and techniques.

Keywords — physical properties, agency, subjective experiences, brain states, future psychophysics, typology

II. Introduction

At the last Fechner Day meeting, Hubbard (2020) suggested several potential directions and agendas for a psychophysics of the future (e.g., involving study of music cognition, analogy and metaphor, neuroscience and embodied cognition, perception of causality, consciousness, spatial biases, individual and species differences, etc.). These suggestions included directions and agendas involving areas that researchers might not initially consider to be within the general domain of psychophysics. However, the justification for many of the suggested directions and agendas involved consideration within those areas of the relationships between the properties of a referent physical stimulus and the properties of a mental representation of that stimulus, and such relationships are, of course, cornerstones of psychophysical theory and investigation. Even so, beyond this general shared consideration, it was not clear how such a range of areas could fit under the general umbrella of psychophysics. Thus, a primary purpose here is to begin to develop a theoretical underpinning for including these additional areas within the broad scope of psychophysics.

Such an endeavor is consistent with the broad consideration of psychophysics in Fechner (1860), Stevens (1975), Baird and Noma (1978), and Kaernbach, Schröger, and Müller (2004), all of whom saw psychophysics as extending beyond measurement of sensation and sensory thresholds. The typology proposed here extends that earlier work by proposing a larger framework within which several different domains of investigation might be considered within psychophysics. Four distinct orders of psychophysics are proposed, and characteristics and examples of each order are provided. The primary distinction between the different orders of psychophysics involves the types of stimulus information to which subjective experiences or brain states are mapped, with such mappings involving information regarding perceived, remembered, or imputed physical properties of stimuli in the first three orders and information regarding agentive properties of stimuli in the fourth order. Additionally, a domain of medial psychophysics that focuses on relationships of physical stimuli and brain states is proposed. Many areas suggested for a future psychophysics in Hubbard (2020) are consistent with third-order and fourth-order psychophysics, and so the majority of discussion here will be on those orders.

III. Medial and First-Order Psychophysics

One distinction within classical psychophysics is between inner psychophysics and outer psychophysics: Inner psychophysics refers to the relationships between brain states and subjective experiences, and outer psychophysics refers to the relationships between physical stimuli and subjective experiences (Fechner, 1860). Thus, classical psychophysics involves the domains of physical stimuli, brain states, and subjective experience. What is missing from classical psychophysics (but is present in contemporary studies of cognitive neuroscience and embodied cognition) is consideration of the relationships between brain states and physical stimuli, and as brain states can be viewed as intermediate between a physical stimulus and the subjective experience, this would constitute a *medial psychophysics*. Both medial psychophysics and outer psychophysics involve relationships of some aspect of the observer (brain states and subjective experiences, respectively) with perceived properties of physical stimuli. Also, we might consider a psychophysics in which information regarding physical properties of a stimulus is present and perceived at the time of an observer's response or judgment to provide the strongest, clearest, and most direct information about that stimulus, and so relationships of brain states or subjective experiences to physical properties of a concurrently perceived stimulus would constitute a *first-order psychophysics*. Much of traditional perceptual psychophysics is of this type (e.g., Gescheider, 2016; Kingdom & Prins, 2016).

IV. Second-Order Psychophysics

As research in the 1970s and 1980s clearly demonstrated, psychophysical techniques and methods initially developed for the study of perception and perceived magnitude could be adapted for the study of memory and remembered magnitude. This has been referred to as *memory psychophysics* (Algom, 1992; Hubbard, 1994). In studies of memory psychophysics, the physical stimulus is not present at the time of the response, and as the response was based on memory rather than on perception, this would constitute a *second-order psychophysics*. Studies in memory psychophysics found that the relationships between remembered magnitude and physical intensity were often well-described by power functions, just as in perceptual psychophysics the relationships between perceived magnitude and physical intensity were often well-described by power functions. However, the exponent of the power function for a given stimulus dimension was usually different in memory psychophysics from the exponent in perceptual psychophysics (i.e., different in second-order psychophysics than in first-order psychophysics). Many other findings typical of perceptual psychophysics such as the symbolic distance effect and the semantic congruity effect were also found to occur in memory psychophysics (e.g., Petrusic, Baranski, & Kennedy, 1998).

One early idea in memory psychophysics literature was the reperceptual hypothesis (e.g., Kerst & Howard, 1978), which suggested that the remembered magnitude of a stimulus in a given dimension was related to the physical intensity of that stimulus by a power function with an exponent approximately equal to the square of the exponent for the perceived magnitude of that dimension (e.g., the exponent relating perceived area to physical area is approximately 0.8, and the exponent relating remembered area to physical area is approximately 0.64 [i.e., 0.8 x 0.8]). The reperceptual hypothesis suggests retrieval from memory is mathematically equivalent to a re-perception of the physical stimulus, but subsequent findings were inconsistent with this suggestion (e.g., Petrusic et al., 1998; Sarfaty, 1986). Also, the magnitude of the exponent for remembered magnitude for a given stimulus dimension could change over time, and the rate and magnitude of change could vary across different stimulus dimensions (Hubbard, 1994). Psychophysical techniques could also be applied to studying properties of mental imagery that did not necessarily involve memory of a specific object that had been previously perceived (e.g., Baird & Hubbard, 1992; Hubbard & Baird, 1988, 1993). To the extent that second-order psychophysics draws on specific remembered exemplars, it would be related to episodic memory.

V. THIRD-ORDER PSYCHOPHYSICS

In second-order psychophysics, the stimulus being judged was previously present and perceived by the observer. However, effects of the likely properties of a physical object on a subsequently perceived stimulus can occur even when those likely properties were never actually physically present. An example of this would be subjective contours of the type made famous by Kanizsa (1979) and illustrated in Figure 1. The configuration of stimuli suggests boundaries where no boundaries physically exist, and observers' perceptual systems insert those boundaries (i.e., observers report a white triangle above a black outline triangle and whose corners are resting on three black circles). Furthermore, the strength of such inserted boundaries is increased as more configural information suggests such boundaries should be present (Schiffman, 2001). Such a case, in which properties of a physical stimulus (e.g., boundaries) are imputed to a perceived stimulus but are or were never actually physically presented, would constitute a *third-order psychophysics*. Such imputation of information in object and scene perception is consistent with the application of the Gestalt principles of perceptual grouping (Wagemans et al., 2012), as well as with many spatial biases (e.g., representational momentum, in which represented location of a moving target is biased in the direction of anticipated motion, Hubbard, 2005, and which could reflect imputation of momentum to the target), and potentially with other examples of "unconscious inference" in perception (e.g., Helmholtz, 1867; Rock, 1983).

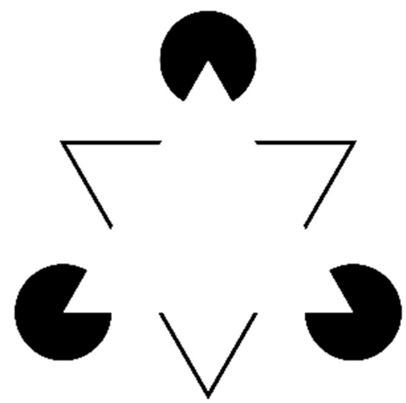


Figure 1. Subjective contours. Observers typically report seeing a white triangle superimposed on a black outline triangle and three circles. Adapted from Kanizsa (1979).

Other examples from cognitive psychology could be framed as third-order psychophysics. One example mentioned in Hubbard (2020) involves perception of (mechanical) causality (also referred to as *phenomenal causality*). A well-known example is Michotte's launching effect, in which causality is not actually present in the stimulus but is nonetheless perceived by the observer (Hubbard, 2013a,b). As illustrated in Figure 2, in a launching effect, a moving object contacts a previously stationary object, and at the moment of contact, the stationary object begins to move; observers often spontaneously report perceiving that contact from the first object caused the second object to move. In the case of an animated display, there is no actual causality (as a computer-animated stimulus does not actually possess mass or force), yet causality is imputed to the first object. A second example involves cognitive simulation of physical systems as reported in the naïve physics literature (e.g., Gentner & Stevens, 1983; Hegarty, 2004; Schwartz & Black, 1999). A third example involves dream experience during REM sleep and delusions or hallucinations during waking. Although qualities of the former might be investigated in real time (e.g., by lucid dreamers, Kahan & LaBerge, 1994), qualities of the latter might be more difficult to investigate in real time, although reports collected after the experience might be possible. To the extent that third-order psychophysics draws on abstracted information not present or previously perceived in specific exemplars, it would be related to semantic memory.

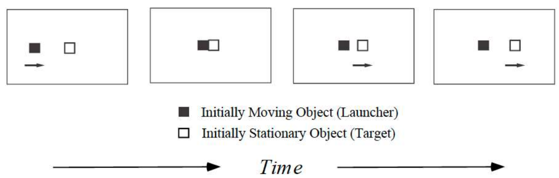


Figure 2. The launching effect. An initially moving object (launcher) contacts a stationary object (target), which then begins moving in the same direction and at the same or a slower velocity. Observers spontaneously report the launcher caused subsequent motion of the target. Adapted from Hubbard (2013a).

An interesting parallel can be seen between the notion of third-order psychophysics and the notion of isomorphism in Gestalt psychology. The notion of isomorphism suggests a brain field or pattern of activity resembles the original stimulus (Koffka, 1935; Wertheimer, 1912). Although the Gestalt view did not suggest a literal structural isomorphism (e.g., an image of a green elephant would involve an area of the brain that was green and shaped like an elephant), but rather suggested a functional isomorphism, the view was often misunderstood (Henle, 1984). Shepard (1975; Shepard & Chipman, 1970) distinguished between structural isomorphism and functional isomorphism, and he referred to the former as *first-order isomorphism* and to the latter as *second-order isomorphism*. As shown in Figure 3, Shepard (1981) used the idea of second-order isomorphism to demonstrate how mental imagery could preserve geometric and kinematic properties of imaged objects (e.g., just as an object rotating from orientation A to orientation C had to pass through intermediate orientation B, so too did the mental representation of a rotating object pass through an intermediate orientation B), and Hubbard (2006, 2019) further developed the notion of second-order isomorphism to include dynamic properties of represented objects (e.g., just as an object rotating from orientation A to orientation C possessed momentum, so too did the mental representation of a rotating object possess [representational] momentum). Second-order isomorphism and third-order psychophysics preserve functional properties of objects, and properties of higher orders of isomorphism appear consistent with properties of higher orders of psychophysics.

VI. FOURTH-ORDER PSYCHOPHYSICS

Thurstone (1927a,b), Stevens (1975), and others have shown how psychophysical techniques can be applied to non-sensory stimuli. In many of these cases, properties regarding the agency of the stimulus, rather than of the physical properties of the stimulus, seem to be the qualities upon which subjective experiences are (and brain states can potentially be) mapped. Rather than subjective experiences or brain states involving representations of physical properties of stimuli, subjective experiences or brain states would involve representations of agentive properties of stimuli, and this would constitute a *fourth-order psychophysics*. Rather than involving physical properties of objects, a fourth-order psychophysics involves agency and intentionality of entities or events. As shown by consideration of first-, second-, and third-order psychophysics, in the higher orders of psychophysics, experiences or judgments are increasingly abstracted away from currently available information regarding the physical properties of stimuli, and in fourth-order psychophysics, information regarding physical properties is replaced by a consideration of the agentive properties of the stimulus.

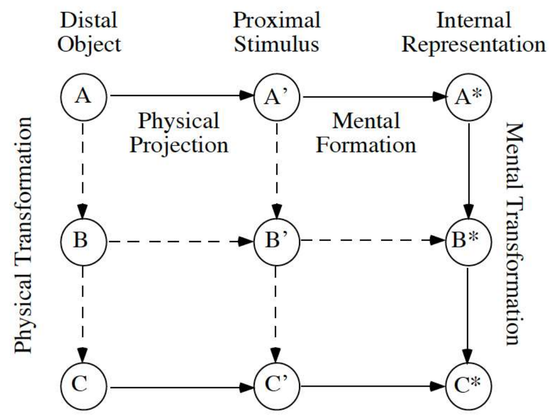


Figure 3. Second-order isomorphism. Just as a distal object undergoing a physical transformation (e.g., rotation) passes through intermediate states, so too does an internal mental representation undergoing transformation pass through intermediate states. Adapted from Shepard (1981).

In many cases, the agentive stimulus or quality to be studied would involve humans or human activity (e.g., scaling the seriousness of crimes and severity of punishments, Gescheider, Catlin, & Fontana, 1982). Even so, it is not the case that fourth-order psychophysics is limited to human agency, as agency can potentially be attributed to any type of stimulus. An early example of this involved the research of Heider and Simmel (1944), in which social motivations and goals were attributed to simple moving geometric shapes. Relatedly, researchers have also attempted to determine the properties of animate stimuli (e.g., Rakison & Poulin-Dubois, 2001). More broadly, literature on social phenomenal causality, in which properties of animacy and agency (e.g., Rutherford & Kuhmeier, 2013), as well as intentions (e.g., chasing, Gao, Baker, Tang, Xu, & Tenenbaum, 2019), are assigned to other stimuli could be considered as fourth-order psychophysics. Attributions of agency can be extended to elements of the natural world, and this is consistent with use of an intentional stance (Dennett, 1987) and social attribution in explaining behaviors of elements of the natural world that is found in shamanism (Hubbard, 2012). Indeed, such a social psychophysics is a promising area of future study (e.g., Anderson, 1982; Jack & Schyns, 2017). To the extent that fourth-order psychophysics draws on information related to animacy and agency, it would be related to social cognition.

The usefulness of fourth-order psychophysics will ultimately be tied to the success (or lack of success) of the reductionist agenda in science, which seeks to explain principles and findings of one science in terms of the principles and findings of a more basic science (e.g., explain psychology in terms of biology, biology in terms of chemistry, chemistry in terms of physics, etc.). In a reductionist account, all phenomena are naturalized (cf. the notion of "supernormal", Radin, 2013) and all explanatory arrows point toward physics (Weinberg, 1994). To the extent that a reductionist approach is not able to discover bridge laws that directly link psychological principles to biological, chemical, or physical principles (e.g., if there are emergent properties at the psychological level that could not be explained by consideration of individual physical components in isolation; Fodor, 1974; van Gulick, 2001; Kauffman, 2008), then reductionism would not be a viable scientific strategy and all explanatory arrows would not point toward physics. If the reductionist program fails, then fourth-order psychophysics becomes more important, as it would offer the most useful level of description, generalization, and explanation of findings in psychology, economics, history, and other less basic sciences.

VII. SIMILARITIES AND DIFFERENCES

There are similarities and differences across the different orders of psychophysics, and these are summarized in Table 1. One similarity involves the range of phenomena accessible to each order, with the possibility of an inner, medial, and outer psychophysics in each order. First-, second-, and third-order psychophysics link measures of the perceived, remembered, and imputed physical properties, respectively, of stimuli with measures of subjective experiences (outer psychophysics) and brain states (medial psychophysics) and link measurements of brain states with subjective experiences (inner psychophysics). Fourth-order psychophysics links measures of the agentive properties of stimuli with measures of subjective experiences (outer psychophysics) and brain states (medial psychophysics) and links measures of brain states with subjective experiences (inner psychophysics). A second similarity is the possibility of perception-action interaction, with action-specific perception occurring within each order, in which perceived, remembered, or imputed physical properties or agentive properties of the stimulus are influenced by the likely actions of the observer.

Table 1.
Comparisons of the Different Orders

Characteristics	First-Order	Second-Order	Third-Order	Fourth-Order
Type of Information	Perceived properties and objects	Remembered properties and objects	Imputed properties and objects	Perceived, remembered, and imputed agency
Distal Properties	Physical	Physical	Physical	Agency, Intention
Labels	Perceptual psychophysics	Memory psychophysics	Top-down processing; unconscious inference	Social psychophysics
Related Domain	Sensory functioning	Episodic memory	Semantic memory	Social cognition
Examples	Weber's law, Stevens' law, sensory thresholds	Reperceptual hypothesis	Stimulus completion, phenomenal causality, representational momentum	Social attribution theory, social causality (chasing, threatening)
Reductionism	More likely	Possible	Less likely	Unlikely
Type of Stimulus	More prothetic than metathetic	More prothetic than metathetic	More metathetic than prothetic	More metathetic than prothetic
Range of Phenomena	Inner, medial, outer psychophysics	Inner, medial, outer psychophysics	Inner, medial, outer psychophysics	Inner, medial, outer psychophysics
Interaction	Action-specific perception	Action-specific perception	Action-specific perception	Action-specific perception

One difference across the different orders involves information that subjective experience and brain states are mapped onto, with first-, second-, and third-order psychophysics involving perceived, remembered, and imputed physical properties of the stimulus, respectively, and fourth-order psychophysics involving agentive properties of the stimulus. The domains related to the different orders are different, with first-, second-, third-, and fourth-order psychophysics related to sensory functioning, episodic memory, semantic memory, and social cognition, respectively. The likely success of reductionism is decreased with higher orders. There might be differences in the frequency of use of specific techniques in specific orders (e.g., quantitative judgments and qualitative judgments might be more common in first-order psychophysics and in third-order psychophysics, respectively). Such differences might be related to whether the stimulus is prothetic (i.e., "how much", e.g., loudness, size) or metathetic (i.e., "what kind", e.g., causality, color); although examples of prothetic or metathetic dimensions might exist in all orders, initial consideration suggests prothetic dimensions might be more common in lower orders and metathetic dimensions might be more common in higher orders.

VIII. SUMMARY AND CONCLUSIONS

A new typology involving four distinct orders of psychophysics was proposed, and the different orders involved differences in whether subjective experiences or brain states were mapped onto perceived, remembered, or imputed physical properties of a stimulus or onto agentive properties of a stimulus. These different psychophysical orders overlap with different domains of psychological investigation, with first-, second-, third-, and fourth-order psychophysics addressing questions typically involving sensory functioning, episodic memory, semantic memory, and social cognition, respectively. Treating agency as a dimension akin to physical properties allows application of psychophysical methods and techniques to social stimuli, but usefulness of fourth-order psychophysics depends on whether such qualities can be reduced to more basic sciences or are non-reducible; in the latter case, a fourth-order psychophysics might yield more useful generalizations and explanations than would reductionist methods. Higher orders of psychophysics might be consistent with higher orders of isomorphism, and the proposed typology could have implications for Gestalt theory. A medial psychophysics focused on the relationships of physical stimuli and brain states was proposed, and this complement focus on the relationships of physical stimuli and subjective experiences in outer psychophysics and focus on the relationships of brain states and subjective experiences in inner psychophysics.

What would be the potential use of such a new typology for psychophysics? One possibility is the presence of a common set of methodologies and techniques. Although in some cases it might be more challenging to collect data for higher-order cases (e.g., needing lucid dreamers to report on intensities and magnitudes experienced during their dreams), in other cases data collection could be relatively straightforward (e.g., studies of naïve physics and of the perception of causality). More importantly, such a typology suggests new methods for studying problems (e.g., methods developed for the study of perception can be applied to the study of memory or social cognition). A second, albeit related, possibility is extending the potential range of psychophysics beyond the traditional studies of threshold and scaling (e.g., subjective contours as an example of third-order psychophysics), and this could help bring phenomena initially seen as unrelated within a common theoretical framework. Examining similarities and differences in the methods, theories, and topics across different orders would provide a theoretical underpinning and constitute a research agenda for psychophysics in the future. Across all orders, though, the goal of psychophysics remains unchanged: an understanding of the relationship between the properties of a stimulus and the representation of that stimulus.

References

Algom, D. (1992). Memory psychophysics: An examination of its perceptual and cognitive prospects. In D. Algom (Ed.). *Psychophysical approaches to cognition* (pp. 441-513). Elsevier/North-Holland.

Anderson, N. H. (1982). Cognitive algebra and social psychophysics. In B. Wegener (Eds.) *Social attitudes and psychological measurement* (pp. 123-148). Taylor & Francis.

Baird, J. C., & Hubbard, T. L. (1992). Psychophysics of visual imagery. In D. Algom (Ed.). *Psychophysical approaches to cognition* (pp. 389-440). Elsevier/North-Holland.

Baird, J. C. & Noma, E. (1978). *Fundamentals of scaling and psychophysics*. Wiley.

Dennett, D. (1987). *The intentional stance*. MIT Press.

Fodor, J. (1974). Special sciences (Or: The disunity of science as a working hypothesis). *Synthese*, *28*, 97-115.

Fechner, G. T. (1860). *Elemente der Psychophysik*. Holt, Rinehart and Winston.

Gao, T., Baker, C. L., Tang, N., Xu, H., & Tenenbaum, J. B. (2019). The cognitive architecture of perceived animacy: Intention, Attention, and Memory. *Cognitive Science*, *43*, e12775.

Gentner, D., & Stevens, A. L. (Eds). (1983). *Mental models*. Erlbaum.

Gescheider, G. A. (2016). *Psychophysics: The fundamentals* (3rd ed.). Routledge/Taylor & Francis.

Gescheider, G. A., Catlin, E. C., & Fortana, A. M. (1982). Psychophysical measurement of the judged seriousness of crimes and severity of punishments. *Bulletin of the Psychonomic Society*, *19*, 275-278.

van Gulick, R. (2001). Reduction, emergence, and other recent options on the mind/body problem: A philosophic overview. *Journal of Consciousness Studies*, *8*(9-10), 1-34.

Hegarty, M. (2004). Mechanical reasoning as mental simulation. *TRENDS in Cognitive Sciences*, *8*, 280-285.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology, 57,* 243-259.

Helmholtz, H. L. (1867). *Handbuch der physiologischen Optik*. L. Voss.

Henle, M. (1984). Isomorphism: Setting the record straight. *Psychological Research*, *46*, 317-327.

Hubbard, T. L. (1994). Memory psychophysics. *Psychological Research/Psychologische Forschung, 56,* 237-250.
Hubbard, T. L. (2005). Representational momentum and related displacements in spatial memory: A review of the findings. *Psychonomic Bulletin & Review, 12,* 822-851.
Hubbard, T. L. (2006). Bridging the gap: Possible roles and contributions of representational momentum. *Psicologica, 27,* 1-34.
Hubbard, T. L. (2012). Causal representation and shamanic experience. *Journal of Consciousness Studies*, *19*, 202-228.
Hubbard, T. L. (2013a). Phenomenal causality I: Varieties and variables. *Axiomathes*, *23*, 1-42.
Hubbard, T. L. (2013b). Phenomenal causality II: Integration and implication. *Axiomathes*, *23*, 485-524.
Hubbard, T. L. (2019). Momentum-like effects and the dynamics of perception, cognition, and action. *Attention, Perception, & Psychophysics*, *81*, 2155-2170.
Hubbard, T. L. (2020). The future of psychophysics: A personal view. In J. R. Schoenherr, T. L. Hubbard., W. Stine, & C. Leth-Steensen (Eds.). *Proceedings of the 36th Annual Meeting of the International Society for Psychophysics: Fechner Day 2020* (pp. 28-33), Ottawa, Canada.
Hubbard, T. L., & Baird, J. C. (1988). Overflow, first-sight, and vanishing point distances in visual imagery. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 641-649.
Hubbard, T. L., & Baird, J. C. (1993). The effects of size, clutter, and complexity on the vanishing point distance in visual imagery. *Psychological Research/Psychologische Forschung, 55,* 223-236.
Jack, R. E., & Schyns, P. G. (2017). Toward a social psychophysics of face communication. *Annual Review of Psychology, 68,* 269-297.
Kaernbach, C., Schröger, E., & Müller, F. (2004). *Psychophysics beyond sensation: Laws and invariants of human cognition*. Erlbaum.
Kahan, T. L., & LaBerge, S. (1994). Lucid dreaming as metacognition: Implications for cognitive science. *Consciousness and Cognition*, *3*(2), 246-264.
Kanizsa, G. (1979). *Organization in vision: Essays on Gestalt psychology*. Praeger.
Kauffman, S. A. (2008). *Reinventing the sacred: A new view of science, reason, and religion*. Perseus Books.
Kerst, S. M., & Howard, J. H. (1978). Memory psychophysics for visual area and length. *Memory & Cognition*, *6*, 327-335.
Kingdom, F. A. A., & Prins, N. (2016). *Psychophysics: A practical introduction* (2nd ed.) Elsevier.
Koffka, K. (1935). *Principles of Gestalt Psychology*. Harcourt Brace.
Petrusic, W. M., Baranski, J. V., & Kennedy, R. (1998). Similarity comparisons with remembered and perceived magnitudes: Memory psychophysics and fundamental measurement. *Memory & Cognition, 26*(5), 1041-1055.
Radin, D. (2013). *Supernormal: Science, yoga, and the evidence for extraordinary psychic abilities*. Random House/ Deepak Chopra Books
Rakison, D. H., & Poulin-Dubois, D. (2001). Developmental origin of the animate–inanimate distinction. *Psychological Bulletin, 127*(2), 209-228.
Rock, I. (1983). *The logic of perception*. MIT Press.
Rutherford, M. D., & Kuhmeier, V. A. (Eds.) (2013). *Social perception: Detection and interpretation of animacy, agency, and intention*. MIT Press.
Sarfaty, O. (1986). *Perceptual and memorial psychophysics function in different sensory domains*. Unpublished M. A. Dissertation, Tel Aviv University.
Schwartz, D. L., & Black, T. (1999). Inferences through imagined actions: Knowing by simulated doing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(1), 116–136.
Shepard, R. N. (1975). Form, formation, and transformation of internal representations. In R. L. Solso (Ed.). *Information processing and cognition: The Loyola Symposium* (pp. 87-122). Erlbaum.
Shepard, R. N. (1981). Psychophysical complementarity. In M. Kubovy & J. R. Pomerantz (Eds.). *Perceptual organization* (pp. 279-341). Erlbaum.
Shepard, R. N., & Chipman, S. (1970). Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology*, *1*, 1-17.
Schiffman, H. R. (2001). *Sensation and perception: An integrated approach*. Wiley.
Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. Wiley.
Thurstone, L. L. (1927a). The method of paired comparisons for social values. *Journal of Abnormal and Social Psychology, 21*(4), 384-400.
Thurstone, L. L. (1927b). Three psychophysical laws. *Psychological Review*, *34*, 424-432.
Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin, 138*(6), 1172-1217.
Weinberg, S. (1994). *Dreams of a final theory: The scientist's search for the ultimate laws of nature*. Random House.
Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung [Experimental studies of seeing motion]. *Zeitschrift für Psychologie*, *60*, 321-378.

Numerosity Perception: Investigating the Temporal Dynamics using Non-Symbolic Stimuli

Aslı Bahar İnan[1] and Aslı Kılıç[2]
[1]Department of Psychology, Cankaya University
[2]Department of Psychology, Middle East Technical University
aslibahar@cankaya.edu.tr; askilic@metu.edu.tr

***Abstract*—** Numerosity perception is one of the fundamental topics of psychophysics and cognitive psychology. Studies on numerosity perception use non-symbolic stimuli or number lines. Numerosity can be investigated using either perception or production tasks. For non-symbolic stimuli, in the numerosity perception task, a set of objects like a group of dots are presented to the participants for them to make an estimate about the number of the stimuli. On the other hand, in the numerosity production task, symbolic stimuli like Arabic numerals are presented to the participants and they are required to generate or produce a group of non-symbolic stimuli that corresponds to the given number. The results of numerous studies using non-symbolic stimuli have shown that there is an underestimation of numerosity judgments for the perception tasks of numerosity, while there is an overestimation for the production task. In this study we investigated the temporal dynamics of numerosity perception by using the response-deadline Speed-Accuracy Trade-off Procedure.

***Keywords*—** Numerosity perception, mental number line, approximate number system, temporal dynamics

I Introduction

Numerosity perception is closely related to mental number representation, magnitude perception and mathematical abilities (Crollen et al., 2011; Dehaene et al., 2008; Krueger, 1972). In numerosity perception tasks, non-symbolic stimuli like a collection of dots are presented to the participants and they are required to give a response using symbolic stimuli like Arabic numerals (Crollen et al., 2011; Izard & Dehaene, 2008; Krueger, 1972; Reinert et al., 2019). Approximate number system (ANS) is suggested to be the mechanism for estimating numbers indirectly (Anobile et al., 2016; Dehaene, 2003). ANS is considered to be an innate system (Dehaene, 2011) with evolutionary advantages for human adults, infants and also animals (Anobile et al., 2016; Dehaene, 2003; Whalen et al., 1999). Research findings related to ANS and numerosity perception suggest that humans use a mental number line which is logarithmically scaled (Crollen et al., 2011; Dehaene, 2011; Izard & Dehaene, 2008; Reinert et al., 2019).

In numerosity perception tasks non-symbolic stimuli are to be mapped to symbolic stimuli corresponding to a mapping of the logarithmically scaled mental number line to the actual, linear number line (Dehaene, 2011). Because the mental number line is suggested to be compressed, the subjective number representation corresponds to a smaller number on the linear number line so that this mapping results in systematic underestimation (Crollen et al., 2011; Izard & Dehaene, 2008; Krueger, 1972; Reinert et al., 2019).

However, when feedback is given either as an inducer prior to the trials only once (Izard & Dehaene, 2008) or after every trial (Krueger, 1984; Price et al., 2014) a reduction for the underestimation of numerosity is reported. Using feedback is suggested to calibrate the mental number representation.

In the current study, we investigated the role of feedback by giving feedback after every trial. We also investigated the temporal course of numerosity perception by applying the response-deadline speed–accuracy trade-off (SAT) procedure (Ratcliff, 2006; Ratcliff & McKoon, 2018) in a two-choice numerosity decision task.

II Method

EXPERIMENT 1

Participants. Participants were 10 university students.

Apparatus and Stimuli. The experiment was conducted on PsychoPy2 using the DotStim function (Peirce et al., 2019) with 10-px sized dots presented in a 600 × 600 pixels circle field centered on a 1,280 × 720 pixels monitor. The number of dots presented was sampled randomly from the Binomial distribution with a size of 100 dots and four probability conditions, 0.30, 0.45, 0.55, and 0.70, randomly assigned for each trial.

Design and Procedure. There were three sessions, including four blocks of 140 trials after the practice block in each session. There was a total of 1,680 trials for each participant. Dots which have a range of 10-90 were presented at the center of the screen. A signal that cued the participant to respond at 60, 100, 200, 500, 700, 1,500, 3,000 ms after the stimulus onset. Participants should respond within 500 ms after the signal. Participant's decision was about whether the number of dots presented were greater than 50. If they decided that the number of dots exceeded 50, they pressed "m" else they pressed "z" on the keyboard. After each response, as feedback, they were presented with the number of dots presented. The experiment was a 7 (lag conditions: 60, 100, 200, 500, 700, 1500, and 3000 ms) × 4 (dot condition: 0.30, 0.45, 0.55, 0.70) within-subjects factorial design. There were 60 trials for each condition. The trials were removed from the analysis, if the participants did not respond within 600 ms or if they responded before the signal.

EXPERIMENT 2

Participants. Participants were 11 university students.
Apparatus and Stimuli. They were the same as used in Experiment 1.
Design and Procedure. The difference between Experiment 1 was that the participant were asked to make a decision about whether the number of dots presented were less than 50. To investigate whether asking participants to make their decisions by changing the direction of comparison would change their numerosity judgments.

EXPERIMENT 3

Participants. Participants were 10 university students.
Apparatus and Stimuli. They were the same as used in Experiment 1.
Design and Procedure. The difference between Experiment 1 and 2 was that no feedback was provided.

EXPERIMENT 4

Participants. Participants were 13 university students.
Apparatus and Stimuli. They were the same as used in Experiment 1.
Design and Procedure. The difference between Experiment 1 was that the response keys were reversed. The participants were to respond whether the number of dots presented were greater than 50. If the participants decided that the number of dots exceeded 50, they pressed "z" else they pressed "m" on the keyboard.

III RESULTS

Speed Accuracy Trade-off Functions

The probability to respond "greater than 50" is estimated with an exponential function, which provides independent and unbiased estimates of asymptotic probabilities and processing speed.
The data were fitted with the exponential function
$P\ (Dots > 50) = \lambda_1 + (\lambda_2 - \lambda_1)\ (e - \beta^{(t-\delta)});\ t > \delta;\ \text{else}\ \lambda_2$
using the optim function in R (R Core Team, 2019) to estimate the seven parameters with the maximum likelihood estimation (MLE) method.
P('yes') is the probability to respond that the number of dots being greater than 50; λ_1 is the asymptotic probability to accept the number of dots as being greater than 50 at the late response lags; λ_2 is the asymptotic probability to accept the number of dots as being greater than 50 at the early response lag conditions (at chance level before the information begins to accumulate); β is the rate of accumulation towards the asymptotic probability at later lags; δ is the time point when the information begins to accumulate, which shows the point when the probability departs from chance; and t is the total processing time that includes the time before signal onset and latency.

As the probability of dot sampling increased, more dots exceeded the threshold of 50, and more trials were responded as being greater than 50. Also, independent of the dot condition, at early lags participants responded "greater than 50" more often than they did in later lags. At earliest lags, participants overestimated the number of dots presented. Decisions made at early lags indicate the probability of responding "yes" to the question before the evidence accumulation starts, which is the performance at chance level.

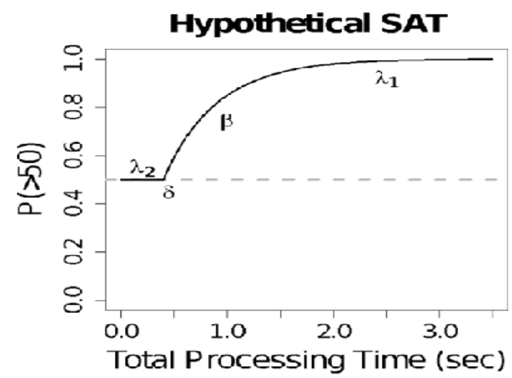


Figure 1. Illustration of hypothetical speed–accuracy trade-off function

IVDISCUSSION

The results from all experiments indicated a tendency towards accepting "greater than 50" response before evidence is accumulated, which is the response bias observed at early lags. The results of all four experiments agreed on an early bias towards responding yes, independent of the differences in the procedures. A general finding from all four experiments was that once accuracy reached its maximum, participants still showed a slight underestimation, even when the number of dots on average exceeded 70.

References

Anobile, G., Cicchini, G. M., & Burr, D. C. (2016). Number as a primaryperceptual attribute: A review. *Perception, 45(1/2)*, 5–31. https://doi.org/10.1177/0301006615602599
Cheyette, S. J., & Piantadosi, S. T. (2019). A primarily serial, foveal accumulator underlies approximate numerical estimation. *Proceedings of the National Academy of Sciences of the United States of America, 116*, 17729–17734.
Crollen, V., Castronovo, J., & Seron, X. (2011). Under- and over-estimation: A bi-directional mapping process between symbolic and non-symbolic representations of number? *Experimental Psychology, 58*, 39–49. https://doi.org/10.1027/1618-3169/a000064
Dehaene, S. (2003). The neural basis of the Weber–Fechner law: A log-arithmic mental number line. *Trends in Cognitive Sciences, 7*, 145147. https://doi.org/10.1016/S1364-6613(03)00055-X
Dehaene, S., Izard, V., Spelke, E.S., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science, 320(5880)*, 1217–1220. https://doi.org/10.1126/science.1156540
Dehaene, S. (2011). *The number sense: How the mind creates mathematics (2nd ed.)*. Oxford University Press.
Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition, 106*, 1221–1247. https://doi.org/10.1016/j.cognition.2007.06.004
Kılıç, A., İnan, A.B. (2021) Response bias in numerosity perception at early judgments and systematic underestimation. Attention, Perception and Psychophysics.

https://doi.org/10.3758/s13414-021-02365-3
Krueger, L. E. (1972). Perceived numerosity. *Perception, & Psychophysics, 11(1)*, 5–9. https://doi.org/10.3758/BF03212674
Krueger, L. E. (1984). Perceived numerosity: A comparison of magnitude production, magnitude estimation, and discrimination judgments. *Perception & Psychophysics, 35(6)*, 536–542. https://doi.org/10.3758/BF03205949
Price, J., Clement, L. M., & Wright, B. J. (2014). The role of feedback and dot presentation format in younger and older adults' number estimation. *Aging, Neuropsychology, and Cognition, 21(1)*, 68–98. https://doi.org/10.1080/13825585.2013.786015
Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman. E., & Lindeløv, J. K. (2019) PsychoPy2: Experiments in behavior made easy. *Behavioral Research Methods, 51*, 195–203. https://doi.org/10.3758/s13428-018-01193-y
R Core Team. (2019). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/
Ratcliff, R. (2006). Modeling response signal and response time data. *Cognitive Psychology, 53(3)*, 195–237. https://doi.org/10.1016/j.cogpsych.2005.10.002
Ratcliff, R., & McKoon, G. (2018). Modeling numerosity representation with an integrated diffusion model. *Psychological Review, 125(2)*, 183–217. https://doi.org/10.1037/rev0000085
Reed, A. V. (1973). Speed-accuracy trade-off in recognition memory. *Science, 181(4099)*, 574-576
Reinert, R. M., Hartmann, M., Huber, S., & Moeller, K. (2019). Unbounded number line estimation as a measure of numerical estimation. *PLOS ONE, 14(3)*. https://doi.org/10.1371/journal.pone.0213102
Stevens S. S. (1959). Cross-modality validation of subjective scales for loudness, vibration, and electric shock. Journal of *Experimental Psychology, 57(4)*, 201–209
Stevens, S. S. (1966). Matching functions between loudness and ten other continua. *Perception & Psychophysics, 1(1)*, 5–8. https://doi.org/10.3758/BF03207813
Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Non-verbal counting in humans: The psychophysics of number representation. *Psychological Science, 10*, 130–137. https://doi.org/10.1111/1467-9280.00120

Evaluating Binary Decisions: Models, Predictors, and Collaboration

Diana Kornbrot
School of Life and Medical Sciences, University of Hertfordshire, UK
d.e.kornbrot@herts.ac.uk

***Abstract*—** In spite of decades of work, analysing the effect of predictor variables on binary decision making is still far from satisfactory. Information accrual models are *broadly* successful in accounting for difficulty, response bias and speed accuracy bias. BUT there is no consensus as to 'best' models. Furthermore, such models are rarely, if ever, reported in a reproducible manner. The *code* to extract parameters is typically not reported in full. Consequently, there is a lack of usable tools., Even when the *raw* data is reported in full, the data underlying tables and graphs is usually absent. Typically, estimates of model parameters for each participant in each condition are not reported. A simpler approach is to use descriptive measures: e.g., error rates, d' and reaction times distribution parameters. This Ms. is a clarion call for a collaborative effort to analyse a small number of data sets, using both specific models and simple descriptive statistics. The aim is to provide easily usable tools for future research. As a spur, a single data set is analysed using the simple descriptive approach. Features emerge that may be missed by information accrual models. Speed and easy conditions are associated with faster, but *less skewed* reaction time distributions than accuracy and hard conditions. Performance depends on both current and previous trial type. Both speed and accuracy are better for correct and previous correct trials. Accuracy depends on reaction time percentile, longest times having worst.

***Keywords*—** binary decisions, response bias, speed bias, information accrual, diffusion decision model

I. INTRODUCTION

Binary decision-making modelling of *accuracy* is one of psychophysics' major successes. Signal detection theory and Luce's choice model enable separate identification of performance ability and response bias. Predictions of these 2 models are indistinguishable in practice. The predominate use in perceptual and memory as well as in applied areas has been to have bias free measures of performance. Sadly, these advances are frequently ignored in applied research where % correct is frequently used to measure performance, Sigh…

Models of reaction time have been less successful in unequivocally identifying measures that are free of speed-accuracy bias. The information accrual parameters of diffusion decision models, (Link, 1975a), (Ratcliff, 1978) DDM and the linear ballistic accumulator, (Brown & Heathcote, 2008), (Heathcote et al., 2002) LBT are candidates, However, parameters are hard to estimate and depend on model details Consequently, they are rarely, if ever, used in applied research. It is hard to know whether speed-based measures of gender or race bias are or are not influenced by speed-accuracy bias.

This work takes a simpler approach and uses d' as a measure of accuracy and mean and skew of reaction time to estimate potential of time criteria. The effect of current and previous trial correctness is considered as well as the directly manipulated predictors.

The intention is to be a prelude to a collaborative analysis project where different team of analysts apply their models to a small number of quality data set

II. A COLLABORATIVE DATA SET

Dutilh et al have produced pioneering work (Dutilh et al., 2018), where a substantial number of research teams analyse the *same* data set. The analyses in this Ms. are based on their data set.

A. Dutilh et al.'s Investigation

1) The Experimental Task and Predictor Manipulations

The task comprised detecting the direction of motion of moving displays. There were 20 participants, all of whom performed in all conditions. The predictors were caution instruction (speed, accuracy), difficulty (easy, hard depending on salience of display motion) and bias via stimulus probability (no, more left moving, more right moving). In addition, analyses were performed according to trial and previous trial type (hit, false alarm (FA) correct reject (CR), miss). Accuracy performance, d', was also assessed as a function of the percentile of reaction time distributions according to condition (accuracy easy, accuracy hard, speed easy, speed hard).

2) Analysts' Tasks and Models/Approaches

Dutilh et al produced 14 subsets of the data each with two conditions A and B. The analysts' task was to decide which of the predictors were different for A and B. For example, A and B might differ *only* as to one predictor – e.g., caution level, or as to two predictors – e.g., caution and bias.

The analysts' models fell into three main classes: diffusion decision model, DDM, (as expounded by Ratcliff); linear ballistic accumulator, LBA; and heuristic. The DM and LBA came in a simple version with no within condition variation in decision parameters and more complex with at least one decision parameter varying within condition (drift rate or starting point or barrier separation). The heuristic analysts used summary statistics of the data, e.g., mean reaction time, proportion of errors, signal detection d'. So, there were 5 model types.

Analysts varied according to the cut-offs used to 'clean' the reaction time, RT, data. There were at least 6 methods: keep all data, remove all data outside N standard deviations from the mean (n =3, 5, or 8), chose arbitrary limits (>200 & < 2000; <150 & <3000, etc.).

All models, of course, assumed that reaction time was the sum of a non-decision time t0 and a decision time that their models modelled. Estimating t0 is controversial and hard.

All models estimate t0 by relating discrimination (d') to properties of the RT *distribution*. Some used moments (mean and variance), mostly LBA analysts; others used percentiles mostly DM. Choice of t0 estimation method is not theoretically tied to the type of DM, LBA.

Three main methods were used to identifying man manipulated predictors. Goodness of fit evaluated model fit with and without a hypothesised predictor. Parameter estimation compared effect of condition using traditional F or t-tests and/or Bayes methods. Eyeball methods were idiosyncratic, based on raw summaries of RT and d'.

3) Dutilh at al.'s Results & Recommendations

As they note core results show considerable consistency as may be seen in Table 5 and Fig. 3. All methods identified where easy instructions had been manipulated, either alone or in combination. Most methods identified where caution had been manipulated, with 3 LBA misses Accuracy conditions had longer means RTs and higher accuracy. Most methods also showed a difference in t0. Manipulating caution often resulted in a difference in ease parameter(s). Authors label this a false alarm. However, in my view it is may be a natural response to caution instructions

They note that LBA slightly underperforms DDM as it 'misses' more effects. Moments is superior to percentiles for estimating t0. There are no other solid findings. The Ms. amply demonstrates the lack of consensus among psychologists, even for simple 2 choice experiments.

4) Summaryof Dutilh eta's findings

This rich data set is amenable to factorial analysis, so that one may identify interactions between manipulated predictors that may be important in practical situations. For example, one might like to know if time pressure has the same effect on easy. Many real-world situations are likely to be non-neutral. In screening applications non target stimuli (benign) may be far more frequent than target stimuli (malignant). Furthermore, missing a malignant stimulus may have far worse consequences than falsely classifying a benign stimulus as malignant.

B. Further Heuristic Analyses: Current Study

The model-based analyses, for reasons of tractability, ignore potentially key features, specifically history (the effect of the previous trial) and the non-normal shape of the RT distributions. This work aims to explore these features.

1) Measures

The discrimination measures used were signal detection d' = z(hit) +z (correct reject) and cut point z(fa). Results are similar for logit(hit) +logit(correct reject as recommended by Link and Heath (Link, 1975b). A correction was made for zero cells, such that if any cell was zero delta was added to that cell and to the corresponding cell. In that event, p(hit) = [n(hit) + delta)]/[n(hit)+ n(miss)+2 delta] and p (correct reject) = [n (correct reject] + delta)]/ [n (correct reject) + n (false alarm) +2 delta]. Value of delta = .5 was used for shown results but using delta = .1 was similar.

The response time (RT) measures used were mean, variance and skew. All measures were estimated for each participant in each cell.

2) Analyses

Five factor factorial ANOVAs were performed on 5 raw measures: mean (RT), variance (RT) skew (RT), d', bias. The predictor factors are caution (speed, accuracy); difficulty (hard, easy); bias (left, no, right), trial type (hit, correct reject. miss, false alarm) and previous trial type. Performance accuracy, d', was examined separately for 4 conditions (accuracy easy, accuracy hard, speed easy, speed hard) using the following percentile categories: 0-25%ile, 25-50%ile, 50-75%ile, 475-90%ile, 75-100%ile.

3) Predictions and explorations

- Speed conditions will be fast BUT inaccurate (low d')
- Easy conditions will be fast BUT accurate (high d')
- Faster conditions will have higher skew
- Errors will differ from corrects on some measures
- Some measures will depend on accuracy of previous trial
- Caution will effect <u>shape of d' v percentile function</u>

C. Results

Full inferential results are available in supplementary materials. All described effects are statistically reliable.

1) Reaction time distributions

Accuracy conditions were slower than speed conditions, and hard conditions were slower than easy conditions.

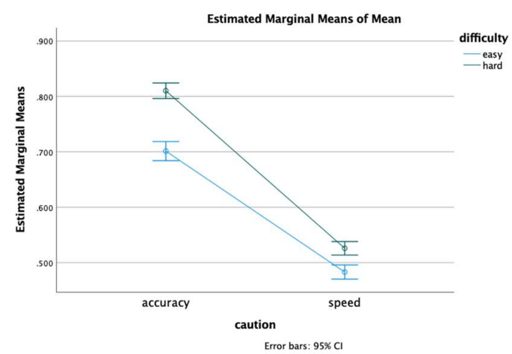


Figure 1. RT mean secs. as function caution & difficulty

By contrast the *slower* accuracy and hard conditions were *more* skewed than the *faster* speed and easy conditions.

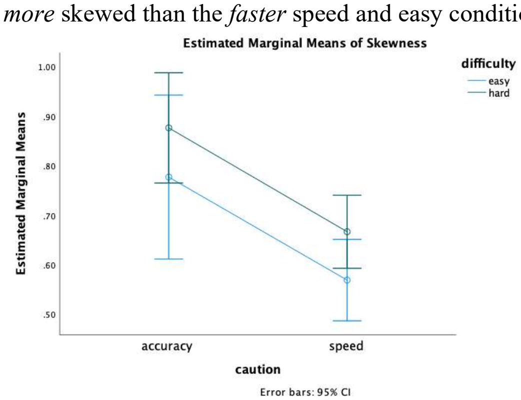


Figure 2. RT skew as function caution & difficulty

Effects of trial type occurred only for accuracy, where errors were *slower* and *less* skewed for errors.

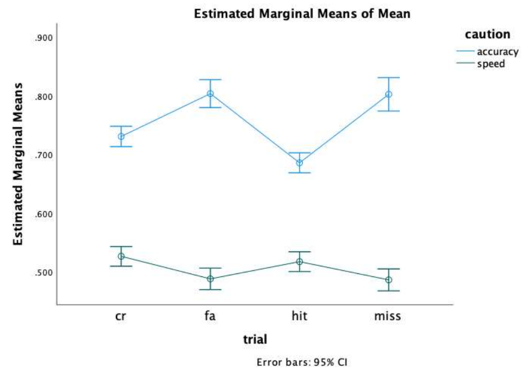


Figure 3. RT mean, secs as function of trial type

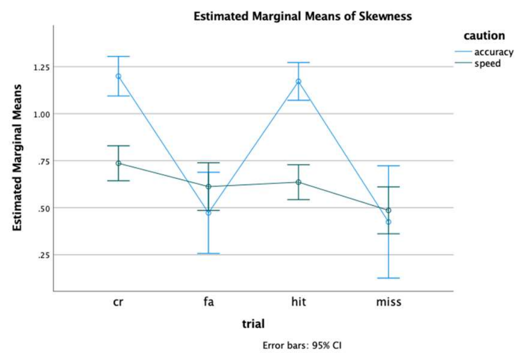


Figure 4. RT skew as function of trial type

Responses were also slower and less skewed following an error, again only for the accuracy condition.

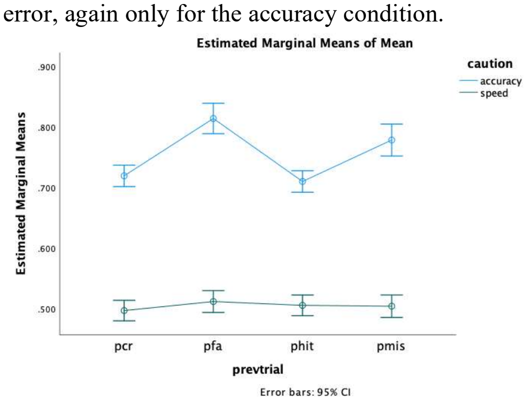


Figure 5. RT mean, secs as function of previous trial type

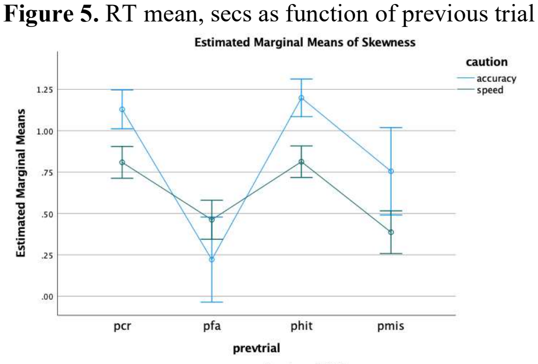


Figure 6. RT skew as function of previous trial type

2) Accuracy and predictors

Accuracy and easy conditions had higher values than speed and hard conditions.

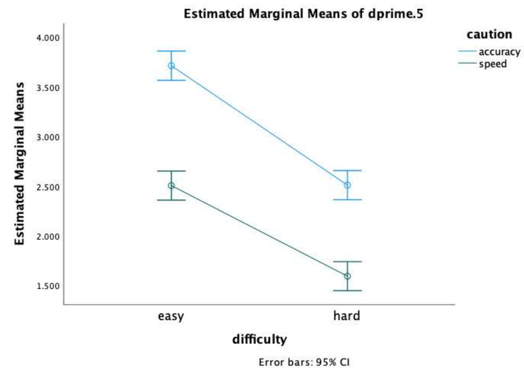


Figure 7. d' as function of caution and difficulty

The value of d' was higher following a correct response this than following an error response. There was an interaction effect such that this effect not apparent in hard conditions.

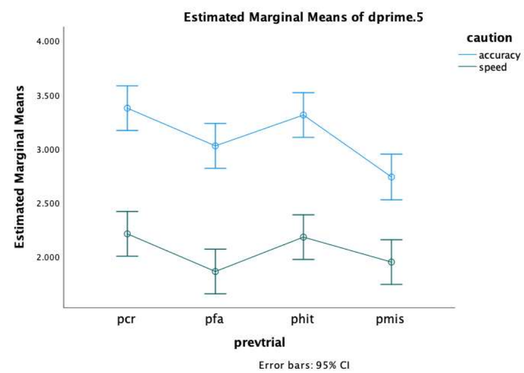


Figure 8. d' as function of trial type

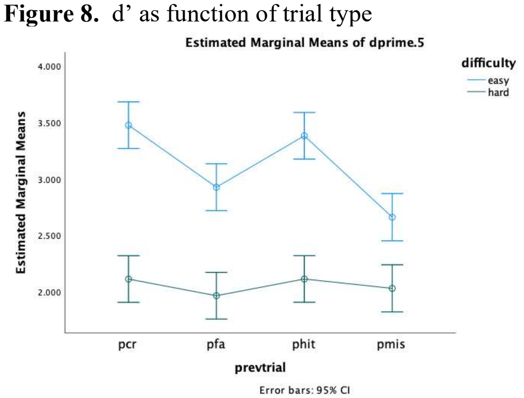


Figure 9. d' as function of previous trail

3) Accuracy as a function of percentiles

The lowest and highest quartiles show the least good performance. Not surprising that short times are less accurate. However very long times are also associated with poor performance. Maybe this is attention lapse

The summary statistics do not give an estimate of non-decision time, but it is notable that accuracy conditions have a mean of minimum over participants of .585 (.574, ,594) seconds while accuracy conditions have mean of .369 (.361,.377) seconds. Simple visual reaction time is of the order of .2 seconds. It appears that speed pressures bring down preparation time, but the relation between percentile and d' is too muddy to infer non-decision time form these summaries. Model analysts mostly found that wherever there are effects on boundary parameters there are also effects on non-decision time.

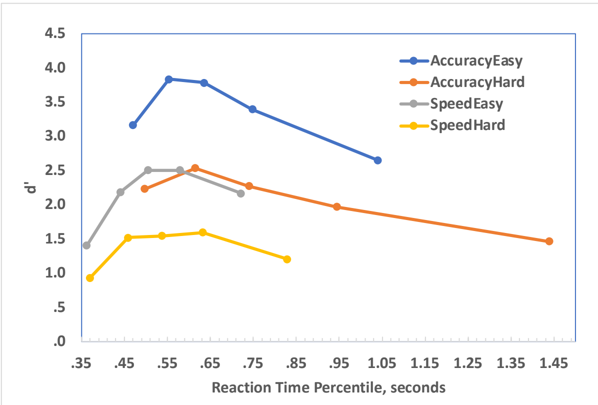


Figure 10. d' as function of mean percentile time, secs.

D. Summary of Results

For the most part these analyses gave standard findings, similar to summaries of either LBA or DDM

There are however some surprises. Manipulating both accuracy and difficulty showed *lower* skew associated with slower responses. This is contrary to the prediction. Another surprise was that trials associated with longer percentiles were less accurate (lower d'). Implications for high accuracy, easy real-world tasks such as spotting malignant cells, may be that regular rest periods are desirable. Implications for task suc as spotting terrorists or identifying implicit se sir gender nias reman it be explored

The findings that correct trials and previous correct trials were associated with better performance for both speed and accuracy, suggesting a higher information accrual parameter is important, but would be hard to obtain from models as there are so few observations in each cell. Here, as predicted, faster responses were more skewed.

There is clearly scope for improving decision making theory, practice and tools.

III. Improving Decision Making & Reproducibility

A collaborative project could very much improve theoretical modelling and practical applications. Many features need attention. It is amazing how many Mss. are unreproducible because of lack of code!

A. Design

Using payoff matrices (possibly convertible to cash) for response and/or caution bias can be more effective than verbal instruction and stimulus frequency. Three conditions are far more informative than 2 conditions and may be enabled by payoff matrices.

B. Deposition of 'raw' data

Should include derived data underpinning tables and graph. This means values of each parameter for each condition for each participant. Authors cannot provide this, even though they *must* have had it to produce the tables and graphs.

C. Analyses deposits

Should include all data cleaning methods, and preferably results with and without cleaning.

Should include R, or other non-proprietary, script to generate raw data underpinning graphs and tables in Ms.

D. Tools

It would be so nice if anyone could put a data set including, for each participant: time and accuracy for each trial and identity of each predictor variable and get out the parameter estimates for each participant in each predictor condition. See example data. Existing tools include (Wagenmakers et al., 2008)(Molenaar et al., 2015) (Lindeloev, n.d.). These tools tend to be for a *single* distribution. So far, I have not found such a tool.

1) GOAL: Tools to get information accrual parameters

Predicting How accuracy, d' depends on reaction time within a condition is very informative

2) GOAL: Tools to plot d' against chosen percentiles

IV. Collaboration

1) Goal identify candidate data sets for collaboration

2) Goal identify Collaborators

Please can anyone who is interested in collaborating: as an analyst, as a tool creator, or as a data provider contact me at d.e.kornbrot@hers.ac.uk

V. References

Dutilh, G., Annis, J., Brown, S. D., Cassey, P., Evans, N. J., Grasman, R. P. P. P., Hawkins, G. E., Heathcote, A., Holmes, W. R., Krypotos, A.-M., Kupitz, C. N., Leite, F. P., Lerche, V., Lin, Y.-S., Logan, G. D., Palmeri, T. J., Starns, J. J., Trueblood, J. S., van Maanen, L., … Donkin, C. (2018). The Quality of Response Time Data Inference: A Blinded, Collaborative Assessment of the Validity of Cognitive Models. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-017-1417-2

Heathcote, A., Brown, S., & Mewhort, D. J. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review*, *9*(2), 394–401. Scopus. https://doi.org/10.3758/BF03196299

Lindeloev. (n.d.). *Reaction Time Distributions: Shiny demo*.

Link, S. W. (1975a). The Relative judgment theory of two-choice reaction time. *Journal of Mathematical Psychology*, *12*, 114–135.

Link, S. W. (1975b). The Relative judgment theory of two-choice reaction time. *Journal of Mathematical Psychology*, *12*, 114–135.

Molenaar, D., Tuerlinckx, F., & Maas, H. L. J. van der. (2015). Fitting Diffusion Item Response Theory Models for Responses and Response Times Using the R Package diffIRT. *Journal of Statistical Software*, *66*(1), 1–34. https://doi.org/10.18637/jss.v066.i04

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Wagenmakers, E.-J., van der Maas, H. L. J., Dolan, C. V., & Grasman, R. P. P. P. (2008). EZ does it! Extensions of the EZ-diffusion model. *Psychonomic Bulletin & Review*, *15*(6), 1229–1235. https://doi.org/10.3758/PBR.15.6.1229

Mentalarchaeology

Stephen Link
Department of Psychology, University of California, San Diego
slink@ucsd.edu

Abstract— The idea of using our modern scientific theories to exam elements of the past is common to many fields. In psychology I know of no such investigations. My purpose is to show how a current psychophysical theory, The Wave Theory of Difference and Similarity, can expose elements of the thought process used by Gustav Fechner when making judgments of weights in 1854-55. Using the data from his carefully designed experiment I will extract measures of decision time, although no response times were made, and measures of the amount of comparative difference used for his judgments.

Keywords— mental archaeology, psychophysics, Wave Theory

The Gain and Loss from Inter-Channel Interactions in the Parallel Systems

Yanjun Liu[1,2] and James T. Townsend[2]
[1]Vanderbilt University
[2]Indiana University Bloomington
yanjun031130@gmail.com; jtownsen@indiana.edu

Abstract— Context effect describing the environmental affects on the interested cognitive processes is an important aspect of cognitive psychology and has been widely studied in perception, and other domains. The violation in marginality is one kind of context effect, which occurs when the marginal processing-time distribution of one channel in a double-channel process is different from that in a single-channel process. In this study, we proposed to use a reaction time (RT) based measurement – workload capacity to examine this empirically unobservable context effect within parallel correlated systems. The workload capacity depicts the change of total processing efficiency with the change of number of internal processes. Our mathematical derivations showed that, if marginals were invariant and internal channels were correlated, the capacity patterns were different across OR and AND stopping-rule conditions, since the correlation only affected the joint but not marginal RT distributions. However, when the marginality was violated, capacities could be similar across various stopping-rules due to trade-offs between affects on marginal distributions and the joint distribution. Our simulation results from two classes of real-time correlated dynamic systems, either assuming marginality or not, supported these theoretical predications. To demonstrate the application of this method, we estimated workload capacities from real empirical data collected from a double factorial dot-detection task and found that some subjects violated the marginality in this experimental setting. Our work provided both theoretical proofs and empirical application to support that workload capacity is a useful tool to examine certain context effect in human's cognitive processes.

Keywords— Parallel processing; channel interactions; correlated channels; race inequality; stochastic linear systems.

VI. INTRODUCTION

Systems involving more than one sub-channels are ubiquitous in perception, decision-making and other domains of cognitive psychology. In a daily scenario where a passenger is about to cross a busy intersection in a big city, he will have to detect both the traffic light and the road condition to ensure that it is safe for him to cross. In this scenario, the cognitive system of the passenger involves two detection processes, each of which is correspondent to a separate visual input, the traffic light and the road condition. In another case, the passenger is at an empty intersection of a small town. Then the passenger may only need to attend to a single detection process for the traffic light to cross the street. The question about how the total processing efficiency of the systems change with the number of internal channels has attracted many cognitive scientists and psychologies.

One broad class of systems people used to study this question is a parallel system where all the internal channels are postulated to process simultaneously. The performance of such a system can be affected by the interaction between its internal sub-processes dramatically. For example, as documented by the famous Stroop effect (Stroop, 1935; Townsend & Algom, 2009), the interaction between the sub-processes impaired the processing efficiency of each other and thus slowed down the total processing efficiency of the system. Contrarily, a positive interaction between the sub-processes can benefit the processing efficiency of each other, and leads to a speedup in the total processing efficiency of the cognitive system as evidenced by the Gestalt effect (Amishav & Kimchi, 2010; Wenger & Townsend, 2001).

In many years, an intensive focus has been devoted to assessing to which extends the strength of these inter-channel interactions can reach in the class of parallel systems. A set of statistical bounds (see details in the next section) for systems following different stopping rules (conjunctive or disjunctive) has been constructed that partially answered this question. A system following a conjunctive (AND) stopping rule does not terminate until the completion of all the internal channels; whereas a system following a disjunctive (OR) stopping rule can terminate after the completion of one internal channel. The derived statistical bounds for the parallel AND systems are the Colonius-Vorberg upper and lower bounds (Colonius, 1990). The derived statistical bounds for the parallel OR systems are the Miller (1982) upper bound and the Grice (1984) lower bound.

The Miller upper bound is also referred as the race inequality. Miller and his colleagues (1982) showed that a coactive system should always violate the race inequality. In a coactive system which is a special class of parallel system, all the information from the parallel processed sub-channels is summed into a single channel before the system terminates. Nevertheless, a common practice that incorrectly inferring the coactivation from observing the violation in the race inequality has been wildly conducted (e.g., Raij, Uutela, & Hari, 2000; Harrar, Harris, & Spence, 2017). In our current investigation, we falsify this common practice by showing that violations in the race inequality can be produced an interactive parallel system and thus cannot suffice the coactivation.

Regardless of the different stopping rules, the abovementioned statistical bounds were derived under the assumption of marginal invariance. A system is said to satisfy the marginal invariance if the marginal performance of its sub-processes is not altered by the inclusion or exclusion of the other channels.

However, in real cognitive scenarios, this presumption may not always be true. To this end, we probe the impact of the sub-channel interaction on the performance of cognitive systems within the broad class of parallel models that either satisfy or violate the marginal invariance.

Extending from the previous theoretical investigations, we examined in the current study that 1) how the total processing efficiency of the parallel systems following different stopping rules can be altered by its internal channels distinctively; 2) how the total processing efficiency of the systems change by their inter-channel interactions at both the joint and the marginal level. In other words, we explored the range of RT behaviors that can be produced by various levels of channel dependence, and with presence or absence of marginal invariance. We will illustrate that RT patterns produced by an interactive parallel race model can also violate race inequality, and demonstrate that, with joint efforts of AND and OR tasks, behavioral patterns of RT can help to detect if there is an effect of dependency on marginal performance of internal cognitive channels.

VII. The Impact of the Inter-channel Interaction at both the Joint and the Marginal Level

Let $F_{AB}(t)$ notate the cumulative distribution function of the total processing time of a parallel system involving channel A and B (double-channel condition). Let $F_A(t)$ and $F_B(t)$ notate the cumulative distribution function of the total processing time of single-channel systems involve either A or B (single-channel condition).

The marginal invariance holds in a parallel system if the processing efficiency of each sub-process within the system is invariant with the alternation in the numbers of the internal channels. Mathematically, it is equivalent to have $P_{AB}(Ta \leq t)\ F_A(t)$ and $P_{AB}(T_B \leq t) = F_B(t)$. Under this assumption, the total processing efficiency of a parallel double-channel system following the AND stopping rule is bounded by its upper and lower bounds (Colonius, 1990) as following:

$F_A(t) + F_B(t) - 1 \leq F_{AB\text{-}AND}(t)$; $F_{AB\text{-}AND}(t) \leq \text{Min}[\ F_A(t), F_B(t)]$.

A parallel double-channel system following the OR stopping rule is bounded by its upper (Miller, 1982) and lower bounds (Grice, 1984): $\text{Max}[\ F_A(t)\ ,\ F_B(t)] \leq F_{AB\text{-}OR}(t)$;
$F_{AB\text{-}OR}(t) \leq F_A(t) + F_B(t)$.

In the current investigation, the inter-channel interaction can impact the performance of the cognitive systems at both the joint and the marginal level. At the joint level, if the inter-channel interaction is positive, then following the probability theory, the processing efficiency of the double-channel system should be better than the products of the marginal performance of its sub-processes. That is, $P_{AB}(T_A \leq t, T_B \leq t) > P_{AB}(T_A \leq t)P_{AB}(T_B \leq t)$. Contrarily, if the inter-channel interact with each other negatively, the joint performance of the double-channel system should be worse than the product of the marginal performance: $P_{AB}(T_A \leq t, T_B \leq t) < P_{AB}(T_A \leq t)P_{AB}(T_B \leq t)$.

At the marginal level, we assume that the marginal performance of the inter-channel processing efficiency should always benefit from the positive interactions between the sub-processes. Thus, the marginal processing efficiency of the internal channels in the double-channel condition should be better than their performance in the single-channel condition. That is, $P_{AB}(T_A \leq t) > P_A(T_A \leq t)$, and $P_{AB}(T_B \leq t) > P_B(T_B \leq t)$. In contrast, the negative interaction between the inter-channels should impair the marginal performance, and therefore the performance of a channel should be worse in the double-channel condition than in its single-channel condition: $P_{AB}(T_A \leq t) < P_A(T_A \leq t)$, and $P_{AB}(T_B \leq t) < P_B(T_B \leq t)$.

VIII. The Impact of the Inter-channel Interaction on Parallel systems followin Different stopping rules

For a parallel system following the AND stopping rule, the inter-channel interaction mainly affects its total processing efficiency at the joint level, as $F_{AB\text{-}AND}(t) = P_{AB}(T_A \leq t, T_B \leq t)$. Thus, a positive inter-channel interaction should always improve the total processing efficiency of the system, while a negative inter-channel interaction should always impair the total processing efficiency of the system. In addition, if the system satisfies the marginal invariance, the gain and loss the system have from its inter-channel interaction cannot go beyond its statistical bounds; whereas if the system allows marginal variability, then its gain or loss from the inter-channel interaction should violate the statistical bounds.

For a parallel system following the OR stopping rule, the inter-channel interaction can affect its total processing efficiency at both the joint and the marginal level. This is because that the stochastic performance of the double-channel OR system can be decomposed as the summation of its marginal performance and the joint performance: $F_{AB\text{-}OR}(t) = P_{AB\text{-}OR}(T_A \leq t \text{ or } T_B \leq t) = P_{AB}(T_A \leq t) + P_{AB}(T_B \leq t) - P_{AB}(T_A \leq t, T_B \leq t)$. $P_{AB}(T_A \leq t) + P_{AB}(T_B \leq t)$ represents the marginal performance if the system, and $P_{AB}(T_A \leq t, T_B \leq t)$ represents the joint performance of the system.

If the parallel-OR system satisfies the marginal invariance assumption, then the marginal performance of each channel in the double-channel condition remains at the same level as in the single-channel condition. Consequently, the positive inter-channel interaction should always impair the total processing efficiency of the system, while the negative inter-channel interaction should always benefit the total processing efficiency of the system. In addition, the gain and the loss the double-channel system can receive from its inter-channel interaction is restricted by the statistical bounds.

On the other hand, if a parallel-OR system violates the marginal invariance, then is inter-channel interaction can affect the performance of the system at both the margional and the joint level. If the internal channels interact with each other in a positive manner, then both the joint and the marginal performance of the system benefit from the interaction. However, due to the tradeoff between the joint and the marginal performance in such a system, the total processing efficiency may either be improved or impaired by its positive inter-channel interaction. If the marginal gain is larger than the joint gain, the total processing efficiency of this system actually is improved by its inter-channel interaction. In contrast if the marginal gain is less than the joint gain, the in fact the system's total processing efficiency is impaired. Contrary patterns of gain and loss are produced by the negative inter-channel interaction. In other words, the total gain or loss in the total processing efficiency of a parallel-OR system violating the marginal invariance is now not only determined by the type of its inter-channel interactions but also the magnitudes of its impact at both the joint and the marginal performance of the system. Subsequently, the total processing efficiency of such a system can be either benefited or impaired by its inter-channel interaction regardless of whether the interaction is positive or negative, and the amount of gain or loss can violate the statistical bounds.

IX. INFERENCES FROM THEORETICAL INVESTIGATIONS

The theoretical investigations presented in the previous section indicate some distinctive patterns of the gain/loss between parallel systems following different stopping rules and with presence or not of the marginal invariance. In this section, based upon these theoretical investigations, we make some theoretical predictions on an RT-based empirical measurement – workload capacity – and show how to use conjoint empirical efforts of different stopping rules to imply the underlying impact of the interactions on marginal and joint performance of the double-channel systems.

A widely used measurement for assessing the change in total processing efficiency in the systems with the alteration in the numbers of the internal channels is the workload capacity (C(t); Townsend & Nozawa, 1995; Townsend & Altieri, 2012). The workload capacity is a ratio that compares the total processing efficiency (as assessed by the cumulative function of its total processing time) of the double-channel systems to the prediction from an unlimited-capacity independent parallel system. If the total processing efficiency of the system is not affected by the change in the number of internal channels, then C(t) is predicted to be 1 and the system is referred to have unlimited capacity. If the system's total processing efficiency improves with the increase in the number of internal channels, then C(t) is predicted to be above 1 and the system has super capacity. Contrarily, if the system is impaired by the inclusion of additional internal channels, then C(t) is predicted to be below 1 and the system has limited capacity. The upper and lower statistical bounds for the parallel systems as introduced in the previous section can be transformed into the C(t) expression that provides clear operational definitions to describe the degree of capacity systems have. If the gain or loss in the system does not violate its upper or lower statistical bounds, then we say the system has moderate super or limited capacity. If the statistical bounds are violated, the system is said to have extremely super or limited capacity, since the gain or loss it has from the change in the number of internal channels exceeds the maximum magnitude predicted by the probability theory.

Now recall the theoretical predictions we have, it implies that if an interactive system follows the marginal invariance, then it should always produce opposite patterns of C(t)s across different stopping rules. In addition, it can never produce C(t)s that violate its statistical bounds. In other words, if there is a positive interaction between the internal channels, the system should always show moderate limited capacity in the OR case, and moderate super capacity in the AND case; whereas if there is a negative inter-channel interaction, such a system should always show moderate super capacity in the OR case, and moderate limited capacity in the AND case.

On the other hand, if an interactive parallel system violates the marginal invariance, then it can produce all types of C(t)s across different stopping rules. That is, the marginal-variable systems cannot only produce the C(t) patterns that mimic those produced by the marginal-invariant systems but also can produce the patterns that can never be observed from the marginal-invariant parallel systems. For example, if the observed C(t)s imply super capacity in the OR case but limited capacity in the AND case, then it immediately falsifies the systems following the marginal invariance and implies that the interaction between the channels affects the system at both the joint and the marginal level. In addition, such a system can also produce C(t) that violates its upper or lower bound in the disjunctive case and thus results in a violation in the race inequality. Our inferences were supported by the simulation results from the class of the linear dynamic systems confirm our theoretical predications and inferences. To save the space, the simulated results will not be presented here but can be find in Townsend, Liu, Wenger and Zhang (2020).

In summary, our investigation showed that under contrasting assumptions of marginal invariance and marginal variability, predictions for the patterns of workload capacity across different stopping rules are dramatically distinct. Moreover, we showed that a parallel race model allowing marginal variability can violate the statistical bounds. Thus, violations of the race inequalities cannot assure coactivation. Last but not the least, combining efforts of AND and OR tasks can help us to test marginal invariance.

References

Amishav, R., & Kimchi, R. (2010). Perceptual integrality of
componential and configural information in faces. *Psychonomic Bulletin & Review*, 17(5), 743-748.

Colonius, H. (1990). Possibly dependent probability summatio
of reaction time. *Journal of Mathematical Psychology*, 34(3), 253–275.

Eidels, A., Townsend, J. T., & Algom, D. (2010). Comparing
perception of Stroop stimuli in focused versus divided attention paradigms: Evidence for dramatic processing differences. *Cognition*, 114(2), 129-150.

Grice, G. R., Canham, L., & Boroughs, J. M. (1984). Combination rule for redundant information in reaction time tasks with divided attention. *Perception & Psychophysics,* 35(5), 451–463.

Harrar, V., Harris, L. R., & Spence, C. (2017). Multisensory
integration is independent of perceived simultaneity. Experimental Brain Research, 235(3), 763–775.

Miller, J. (1982). Divided attention: Evidence for coactivation

with redundant signals. *Cognitive Psychology*, 14(2), 247–279.
Raij, T., Uutela, K., & Hari, R. (2000). Audiovisual integration of letters in the human brain. Neuron, 28(2), 617– 625.
Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643.
Townsend, J. T., & Altieri, N. (2012). An accuracy–response
time capacity assessment function that measures performance against standard parallel predictions. *Psychological Review*, 119(3), 500.
Townsend, J. T., Liu, Y., Zhang, R., & Wenger, M. J. (2020).
Interactive parallel models: No Virginia, violation of Miller's race inequality does not imply coactivation and yes Virginia, context invariance is testable. *The Quantitative Methods for Psychology*, 16(2), 192-212.
Wenger, M. J., & Townsend, J. T. (2001). Faces as Gestalt stimuli: Process characteristics.

The Role of Afterimages in Motion Induced Blindness

Sofia C. Lombardo,[1] Dan Swift,[2] Erika S. Wells,[3] John E. Sparrow,[4] Rebecca A. White,[1] Joseph A. LaBarre,[1] Andrew J. Kitt,[1] and Wm Wren Stine[1]
[1]Department of Psychology, University of New Hampshire, Durham, NH, USA
[2]Department of Psychology, University of Michigan-Dearborn, Dearborn, MI, USA
[3]Department of Psychological and Brain Sciences, Boston University, Boston, MA, USA
[4]Social Science Division, University of New Hampshire-Manchester, Manchester, NH, USA
sofia.lombardo@unh.edu; bill.stine@unh.edu

Abstract— Motion induced blindness (MIB) refers to the perceptual disappearance of a stationary stimulus in the presence of a motion mask. The current study investigated the degree to which afterimages affect MIB inhibition when measured as a contrast detection threshold in a modified replication of White et al. (2020). Adult participants with normal or corrected-to-normal eyesight completed a series of target detection tasks while viewing a standard MIB stimulus with the motion mask removed that consisted of increment versus decrement inducer and target components. A univariate ANOVA data analysis procedure revealed a significant afterimage effect (Scheffé $p < 0.0253$) on contrast detection threshold was found for targets presented at an interstimulus interval of 500 ms. This effect was stronger for decrement targets compared to increment targets in the decrement inducer conditions. Based on a comparison with previous research in which the MIB effect was found to endure across interstimulus intervals up to 15500 ms, the current findings indicate that afterimages do not significantly influence contrast detection thresholds for MIB. Further research is necessary for determining the strength and duration of afterimage effects on contrast detection thresholds in MIB that may be caused by interaction with the motion mask.

Keywords— Psychology, vision, motion induced blindness

I. INTRODUCTION

An effective and commonly employed mechanism for exploring the neuronal and physiological correlates of visual processing is to investigate sensory and perceptual disruptions in which visual stimuli are physically present in the environment but evade conscious awareness (Hofstoetter et al., 2004; Wallis & Arnold, 2008). Motion-induced blindness is a phenomenon in which stationery salient stimuli spontaneously and intermittently disappear and reappear from awareness when surrounded by a pattern of moving distractors (Grindley & Townsend, 1965; Bonneh et al., 2001). MIB is a particularly interesting way through which to investigate visual perception because it represents an example of visual disappearance under natural conditions in normal sighted observers (Bonneh et al., 2001).

The first formal discovery of motion-induced blindness emerged incidentally from Grindley and Townshend's (1965) observation of the conditions of binocular fusion and rivalry in which the positional adjustment of an object in one field of a stereoscope or the occlusion of one binocular field by a moving object occasionally caused the flickering whole or partial disappearance of other objects in the corresponding area of the opposing eye or second binocular field.

There currently is not a widely-accepted consensus on the mechanisms believed to be facilitating MIB, but researchers have proposed a variety of explanations for the phenomenon. Such processes include perceptual filling-in (Hsu et al., 2006), adaptation of target borders (Hsu et al., 2006; Kawabe & Miura, 2007), modulation by occlusion cues (Graf et al., 2002), interhemispheric switching (Carter & Pettigrew, 2003), gain control reduction with response bias shift (Caetta et al., 2007), adaptation and prolonged inhibition (Gorea and Caetta, 2009), and attentional competition between stimulus components (Bonneh et al., 2001).

MIB is often paralleled with a process of surface completion, known as perceptual filling-in (Hsu et al., 2006) or the "Troxler effect," which occurs when an object in the visual field disappears from awareness via displacement by the surrounding background (Hsieh and Tse, 2009). A form of perceptual filling-in constantly occurs in the human visual system when the brain substitutes the blind spot with what would be expected to complete the scene. Because perceptual filling-in necessitates both signal loss of a target stimulus and replacement by the background, it can be conjectured that retinal activation is induced during neuronal adaptation and cortical activation is induced during filling-in (Hsieh and Tse, 2009). The relationship between MIB and perceptual filling-in has been countered by Bonneh et al.'s (2001) claim that early sensory adaptation does not align with two well-known parameters of MIB: slowly moving targets are still susceptible to disappearance and that the effect is regulated by perceptual grouping of MIB targets.

MIB is related to object rivalry, visual field anisotropy, and Gestalt perceptual grouping effects, which are reminiscent of other manifestations of visual disappearance as well as clinical instances of attention deficits in individuals with intact primary visual regions (Bonneh et al., 2001). Attentional theories generally assume that the motion mask attracts attention away from the stationary objects and that perceptual resources are expended on the dynamic stimuli to the point that the static stimuli are consciously ignored. This is presumably an evolutionary trait in which the visual system prioritizes objects in motion over objects in place to maximize available attentional capacity by diverting focus toward potential threats and away from benign entities.

Neuroimaging techniques have identified several neurophysiological pathways associated with MIB and have demonstrated that MIB has distinct neural signatures (Donner et al., 2008), indicating that the spontaneous suppression of the target representation in the ventral pathway is at least partially induced by the mask representation in the dorsal pathway. So MIB appears to be functioning through the inhibition of the ventral stream by the dorsal stream. In addition, White et al. (2020) suggested that MIB occurs within ON/OFF channels in the visual system rather than between channels. ON-center ganglion cells are tuned to detect positive contrast while OFF-center ganglion cells are tuned to detect negative contrast.

Target contrast, size, speed, and flicker rate alongside mask contrast, speed, and density as significant variables that influence target MIB (Bonneh et al., 2001). For example, targets of higher contrast disappear most frequently, dynamic targets disappear, targets with good gestalts tend to disappear or combat disappearance entirely as singular units, and targets with greater spatial proximity to the motion mask disappear more easily than those positioned at greater distances (Bonneh et al., 2001). Wallis and Arnold (2008) found that subjective disappearances of the static targets in MIB did not respond to retinal speed but did respond to temporal frequency, with the maximal disappearances arising at approximately 4Hz.

The purpose of the current study was to measure the role of afterimages in White et al. (2020) using a modified version of their procedure. White et al. presented four stationary targets, which they called inducers, with a motion mask – the typical MIB setup. However, during any given trial the inducers were physically removed and, after a delay, a new target was presented briefly. The luminance contrast detection threshold of the new target was measured. They found that inhibition by the motion mask on the inducers created an increased threshold for the new target that lasted for more than 15 s. Of course, an afterimage would have remained after removal of the inducers, thereby influencing the contrast threshold of the new target. We measured the duration of inducer-generated afterimages using target contrast thresholds. It is hypothesized that the afterimages will have increase contrast thresholds for fewer than 15 s.

II. METHODS

Participants. Participants were recruited via flyers posted on the University of New Hampshire campus and other advertisements distributed by the Vision Lab in the Department of Psychology. Three adult participants with normal or corrected-to-normal vision participated in the study. Demographic information was not collected because it was deemed irrelevant to the topic of interest. Subjects received no compensation for their involvement in the study.

Apparatus and Stimuli. Both the apparatus and stimuli were adapted from White et al. (2020). The primary component of the apparatus was a Dell Dimension E521 computer running *VisionWorks* (Swift, Panish, & Hippensteel, 1997) on a Windows XP operating system projecting through a Mage Systems M2ILH4101 monitor. The 800 x 600 pixel monitor had a 120 dots per inch pixel pitch and 120 Hz refresh. It was equipped with a monochrome P46 ultra-short persistence phosphor (yellow-green; CIE x = 0.427, y = 0.543) and a Vision Research Graphics Gray-Scale Expander VW16 to render 15-bit linearized depth. Supplemental components of the apparatus included a second, 21” flat-screen, monitor, a chin-rest stationed at a distance of 1m from the Mage Systems monitor screen, and a keyboard with the 2, 4, 6, and 8 keys removed.

The stimulus depicted a variation of the standard MIB design employed by White et al. (2020), consisting of the central fixation point, four peripheral inducers, and a flashing target appearing in one of the four quadrants. The motion mask of randomly moving squares that was present in White et al. (2020) was removed. The adaptation screen and trial background maintained a 50 cd/m^2 contrast valence. Increment stimuli (positive contrast valence; bright) possessed a luminance of 90 cd/m^2 and decrement stimuli (negative contrast valence; dark) possessed a contrast valence of 10 cd/m^2. The inducers were either increments, decrements, or the same contrast as the background (no inducer) while the targets were either increments or decrements. Target contrast valence within levels varied by trial. The inducers and targets were presented in the four quadrants at four degrees of retinal angle from the central fixation point.

Design. The current study followed a 2 x 3 x 6 between-subjects experimental design in which there were two levels of target valence, three levels of inducer valence, and six inducer-to-target interstimulus intervals. The six experimental conditions were coded with three letter acronyms to indicate the combination of variables describing the condition, with the first letter representing the mask, the second letter representing the inducers, and the third letter representing the targets. Increment stimuli were coded as "B" for "bright" and decrement stimuli were coded as "D" for "dark." The conditions were NBD (no mask, bright inducers, dark targets), NNB (no mask, no inducers, bright targets), NBB (no mask, bright inducers, bright targets), NDB (no mask, dark inducers, bright targets), NDD (no mask, dark inducers, dark targets), and NND (no mask, no inducers, dark targets).

Procedure. The University of New Hampshire Institutional Review Board (IRB) approved the current study preceding its commencement. The primary researcher obtained informed consent from each participant prior to data collection. Each subject received demonstrations for using the computer equipment, instructions regarding their responsibilities for completing the tasks, and a spreadsheet with a randomly ordered sequence of conditions to track their progress. Participants sat in a darkened room with their head stabilized by a chin rest positioned 1m from the monitor. Before beginning the active portion of the task, participants viewed a five-minute adaption screen. Once the adaptation period elapsed, a five second intermediary preceded the first trail of the condition. A high-pitched tone signaled the onset of each trial. Participants fixated on the central dot for the duration of the task. At the start of each trial, the inducers appeared and disappeared after four seconds. After the physical removal of the inducers, successive targets appeared in one of the four quadrants six times per trial at interstimulus intervals of 500 ms, 3500ms, 5500 ms, 9500ms, 12500 ms, and 15500 ms. A tone signaled the appearance of each target, and participants pressed the button on the keyboard number pad that corresponded with the quadrant in which they perceived the target to appear. In the case that participants failed to perceive the target or otherwise were unsure of its location, they were instructed to make a guess. The number of trials per condition varied between 30 and 40 depending on the value necessary for 20 reversals.

III. RESULTS

Data from all three subjects was included in the analysis without any exclusions. Data were analyzed using a univariate ANOVA procedure for a 2 x 3 x 6 repeated measures experimental design. There were two levels of target valence, three levels of inducer valence, and six inducer-to-target interstimulus intervals, yielding a total of 36 conditions. Each of the three subjects completed all 36 conditions, yielding a total of 108 psychometric functions to be analyzed. Cumulative normal distributions were fit to each of the 108 psychometric functions, producing two analytical variables of interest, the log contrast threshold (mean) and function slope (standard deviation). The log contrast threshold and square root of the standard deviation were normally distributed. Analyses were conducted with a Chi-Muller adjusted F value and the confidence level was set as 0.0253 to control the familywise Type I error rate at 0.05 (from the Šidák inequality) given two variables dependent variables.

The preliminary test involved pooling the three-way interaction of inducer contrast valence by target contrast valence by inducer-to-target interstimulus interval. Because the three-way interaction accounted for little variance in the contrast detection threshold ($p = 0.488$), a new ANOVA was run with the three-way interaction removed. As shown in Table 1, subject ($p < 0.001$), inducer contrast valence ($p < 0.001$), inducer-to-target interstimulus interval ($p < 0.001$), inducer contrast valence by target contrast valence ($p < 0.001$), and inducer contrast valence by inducer-to-target interstimulus interval ($p < 0.001$) were significant. The variable of interest was inducer contrast valence by inducer-to-target interstimulus interval because the objective was to measure the effect of the afterimages produced by the inducers on contrast threshold detection for targets in MIB as a function of time elapsed between the offset of the inducers and the appearance of the targets.

A *post-hoc* Scheffé Test was conducted to pinpoint the location of the effects within the significant interaction between inducer contrast valence and inducer-to-target interstimulus interval using individual one-degree-of-freedom multiple comparisons. One significant interaction was found within the inducer-to-target interstimulus interval variations for the 500 ms interstimulus interval versus the five remaining interstimulus intervals by inducer contrast (Scheffé $p < 0.0253$). Another significant interaction was found within the two-way interactions between target contrast valence and the average of increment and decrement inducer conditions (Scheffé $p < 0.0253$). No additional significant effects within the inducer contrast valence by inducer-to-target interstimulus interval interaction were found.

Afterimage Effect at 500 ms. Figure 1 depicts two mean plots graphing log contrast detection threshold by interstimulus interval for increment versus decrement targets within both increment and both decrement inducer conditions. A strong afterimage effect on contrast detection threshold was observed at the 500 ms interstimulus interval for all four non-control experimental conditions: Increment target – decrement inducer ($10^{1.18}$ = 15.14%), decrement target – decrement inducer ($10^{1.48}$ = 30.20%), decrement target – increment inducer ($10^{1.26}$ = 18.20%), increment target – increment inducer ($10^{1.15}$ =

14.13%). As shown by the tapering slope in the plot for each experimental condition, the afterimage effect decayed between 500 and 3500 ms. The precise time of the decay is not discernible from the current data because no interstimulus intervals between 500 and 3500 ms were measured. This indicates that the afterimage only increased the difficulty for perceiving the target that appeared within the closest temporal proximity to the disappearance of the inducers. There was no effect on the contrast detection threshold for the targets presented at the five remaining interstimulus intervals.

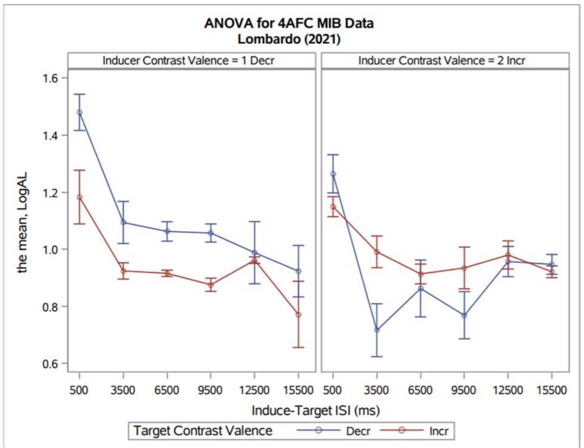


Figure 1. Mean plots of log contrast threshold by inducer-to-target interstimulus interval for increment and decrement targets in decrement inducer and increment inducer conditions.

No Afterimage Effect in Absence of Inducers. Figure 2 depicts the mean plot of log contrast detection threshold by interstimulus interval for increment versus decrement targets within the no inducer condition. Of course, no afterimage effect was observed for the 500 ms interstimulus interval or any of the remaining interstimulus intervals since there were no inducers.

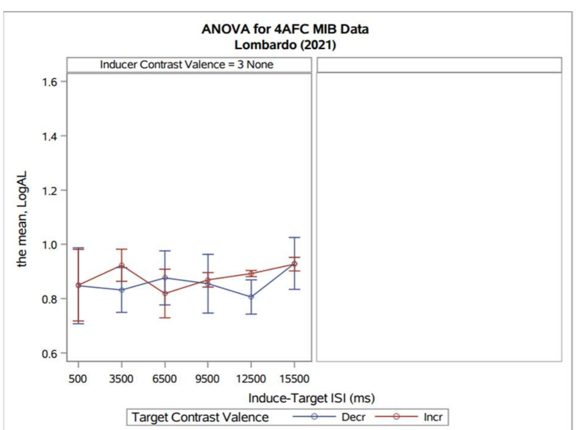


Figure 2. Mean plot of log contrast threshold by inducer-to-target interstimulus interval for increment and decrement targets in no inducer conditions.

Stronger Effect for Decrement Targets with Decrement Inducers. Figure 3 depicts the mean plot graphing log contrast detection threshold averaged across all inducer-to-target interstimulus intervals for increment versus decrement targets by inducer contrast valence. As shown, inducer contrast valence differentially influenced the afterimage effect on target contrast valence. Contrast detection thresholds were higher for both increment (0.95; 0.90) and decrement (1.1; 0.87) targets in the decrement inducer conditions than in the increment inducer conditions. More specifically, a stronger afterimage effect was observed for decrement targets versus increment targets in the decrement inducer conditions.

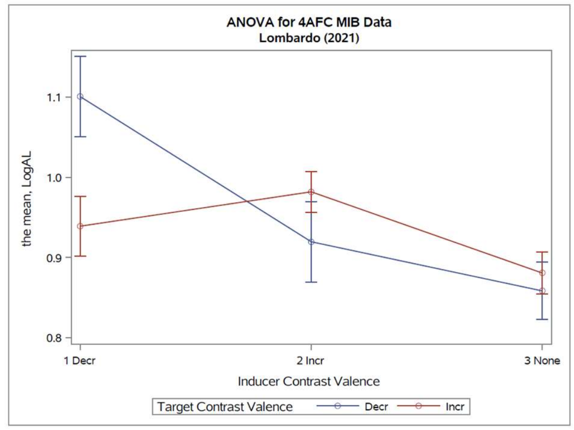


Figure 3. Mean plot of log contrast detection threshold by inducer contrast valence.

IV. DISCUSSION

The primary implication of the results is that afterimages do not significantly impact contrast detection thresholds for MIB. The comparative analysis demonstrates that the MIB effect observed in White et al. (2020) (see Figure 4) and the afterimage effect observed in the current study (see Figure 3) do not align in a way that would indicate that afterimages significantly impact the contrast-dependent detectability of targets by the subject. Although a significant afterimage effect was observed in the current study, it was only present for one of the six interstimulus intervals while the MIB effect in White et al. (2020) endured throughout the six interstimulus intervals. This suggests that the afterimage left by the inducer is brief enough to only impact the perception of the target that appears 500 ms after the disappearance of the inducers and therefore is inconsequential to the MIB effect as a whole. It is evident that afterimages do influence the perception of targets, but it does not seem as though the afterimage effect is modulated through MIB. It is likely that afterimages and MIB do not consistently interact.

The findings extend and replicate several aspects of previous research regarding the relationship between afterimages and MIB. Hofstoetter et al. (2004) found that MIB does not disrupt the formation of afterimages, which indicates that MIB as an illusory phenomenon does not negate afterimages as a pseudo-illusory phenomenon. It is evident from the current study that afterimages can coexist with MIB because the present study demonstrated that the inducers in the MIB paradigm do leave afterimages that can influence how the target is perceived and that this effect overlapped with the MIB effect for the first interstimulus interval. In addition, Hofstoetter et al. (2004) found that afterimage strength does not affect target detection in MIB.

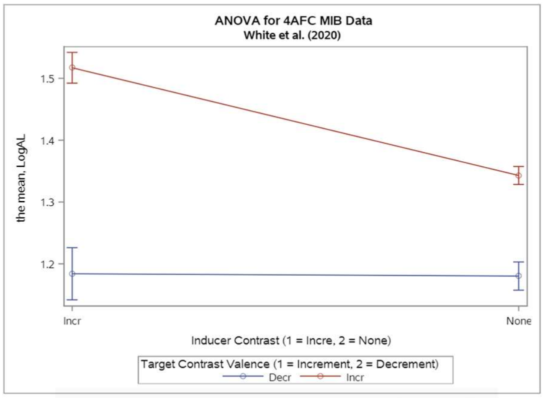


Figure 4. Mean plot of log contrast detection threshold for targets in MIB by inducer contrast from White et al. (2020).

One caveat to the current research is that the data from White et al. (2020) and the current data were collected six months apart, and so the information from both studies may not be comparable given the effect of time on response measures. The next step is to repeat the study using a design that randomly interleaves trials from all conditions in White et al. (2020) and the current study to eliminate any effects from time between tasks. Interleaving condition trials may also correct any testing effects that may have been active in the current study from the sequential repetition of trials within the same condition.

Another avenue for future research is to explore the possibility that the motion mask required for inducing MIB refreshes the afterimage. This would account for both the higher average contrast detection threshold across interstimulus intervals observed in White et al. (2020) and the brief afterimage effect for the first interstimulus interval observed in the current study. It appears that both the afterimage and MIB effects occur simultaneously but that the afterimage effect is much shorter in duration than the MIB effect. The small interval of time in which the two effects overlap is reflected by the heightened contrast detection threshold at the first interstimulus interval. It is difficult to distinguish whether afterimages are inseparable from MIB because the afterimage effects observed in the current study were produced by the inducers, which may be necessary to fully activate MIB. There is evidence, albeit minimal, that there is still an effect on contrast detection threshold for MIB in the absence of inducers, which suggests that MIB may not require inducers but that they enhance the effect. Investigating the prospect that the motion mask in MIB refreshes the inducer afterimages would help clarify the exact role of afterimages in MIB. The challenge for future research would be to determine a suitable methodology for measuring the appearance and effect of afterimages on contrast detection thresholds across interstimulus intervals in the context of MIB.

References

Bonneh, Y. S., Cooperman, A., & Sagi, D. (2001). Motion-induced blindness in normal observers. *Nature*, *411*(6839), 798–801. https://doi.org/10.1038/35081073

Caetta, F., Gorea, A., & Bonneh, Y. (2007). Sensory and decisional factors in motion-induced blindness. *Journal of Vision*, *7*(7), 4. https://doi.org/10.1167/7.7.4

Carter, O. L., & Pettigrew, J. D. (2003). A Common Oscillator for Perceptual Rivalries? *Perception*, *32*(3), 295–305. https://doi.org/10.1068/p3472

Donner, T. H., Sagi, D., Bonneh, Y. S., & Heeger, D. J. (2008). Opposite Neural Signatures of Motion-Induced Blindness in Human Dorsal and Ventral Visual Cortex. *Journal of Neuroscience*, *28*(41), 10298–10310. https://doi.org/10.1523/jneurosci.2371-08.2008

Gorea, A., & Caetta, F. (2009). Adaptation and prolonged inhibition as a main cause of motion-induced blindness. *Journal of Vision*, *9*(6), 16–16. https://doi.org/10.1167/9.6.16

Graf, E. W., Adams, W. J., & Lages, M. (2002). Modulating motion-induced blindness with depth ordering and surface completion. *Vision Research*, *42*(25), 2731–2735. https://doi.org/10.1016/s0042-6989(02)00390-5

Grindley, G. C., & Townsend, V. (1965). Binocular Masking Induced by a Moving Object. *Quarterly Journal of Experimental Psychology*, *17*(2), 97–109. https://doi.org/10.1080/17470216508416418

Hofstoetter, C., Koch, C., & Kiper, D. C. (2004). Motion-induced blindness does not affect the formation of negative afterimages. *Consciousness and Cognition*, *13*(4), 691–708. https://doi.org/10.1016/j.concog.2004.06.007

Hsieh, P.-J., & Tse, P. U. (2009). Microsaccade Rate Varies with Subjective Visibility during Motion-Induced Blindness. *PLoS ONE*, *4*(4). https://doi.org/10.1371/journal.pone.0005163

Hsu, L.-C., Yeh, S.-L., & Kramer, P. (2006). A common mechanism for perceptual filling-in and motion-induced blindness. *Vision Research*, *46*(12), 1973–1981. https://doi.org/10.1016/j.visres.2005.11.004

Kawabe, T., & Miura, K. (2007). Subjective disappearance of a target by flickering flankers. *Vision Research*, *47*(7), 913–918. https://doi.org/10.1016/j.visres.2006.10.026

Swift, D., Panish, S., & Hippensteel, B. (1997). The use of *VisionWorks* in visual psychophysical research. *SpatialVision, 10*, 471-477.

Wallis, T. S., & Arnold, D. H. (2008). Motion-induced blindness is not tuned to retinal speed. *Journal of Vision*, *8*(2), 11. https://doi.org/10.1167/8.2.11

White, R. A., Swift, D., Lombardo, S. C., Wells, E. S., Sparrow, J. E., Oostendorp, A., LaBarre, J. A., Kitt, A. J., & Stine, W. W. (2020). Motion induced blindness: Luminance contrast sensitivity to increments and decrements in the presence of a motion mask. In J. R. Schoenherr, T. Hubbard, W. W. Stine, & C.

Leth-Steensen (Eds.), *Proceedings of the 36th annual meeting of the international society for psychophysics: Fechner day 2020* (pp. 56-59). International Society for Psychophysics.

The Effect of Different Spatial Sound Distributions on Cognitive Load in Audiovisual Perception Tasks

Catarina Mendonça[1,2]
[1]Department of Psychology, University of the Azores, Portugal
[2]Center for Psychology, University of Porto, Portugal
mendonca.catarina@gmail.com

Abstract— In day to day audiovisual applications, sound can occur in varying ranges of space all around the subject, while visual stimuli are limited to the field of view. In two studies, we analyze the impact of sound on visual visual tasks by altering its spatial distribution. In both studies, cognitive effort was monitored through the measurement of pupil diameters. In the first experiment we used immersive 3D movies and manipulated the soundtrack by altering the distribution of sound events. We found higher cognitive effort, quantified in the form of larger pupil diameters, when the sounds were more spatially distributed, compared to more simplified sound conditions. In the second experiment, we used sequences of light flashes and noise bursts with several spatial distributions. Visual localization was affected by some sound distributions, close to the visual distribution, but not by others. Both experiments demonstrate that sound distributions can affect differently the performance in a visual task and that this effect is most likely mediated through changes cognitive load.

Keywords— Audiovisual, Spatial Sound, Cognitive Effort, Cognitive Load, Attention

V. Introduction

Most studies in audiovisual perception are overly simplified and poor at predicting multisensory effects in real-life conditions, where multiple events can occur simultaneously, and attentional factors might play a role in percept formation.

In audiovisual perception, visual stimuli are often the cue with the highest weight, frequently influencing auditory percepts, while sounds rarely yield a relevant influence on visual percepts (e.g., Mendonça, 2020). It could therefore be expected that, in everyday applications, there would be little influence of sounds on our visual perception.

On the other hand, our field of view is limited, and therefore visual stimuli are only presented to us from a narrow spatial area. Conversely, auditory stimuli are presented to a far wider region around us. There is some evidence revealing that sounds can attract visual attention. This attentional guidance might consequently enhance the detection and facilitate the response to visual stimuli within the attended area (e.g. Driver & Spence, 1998; Störmer, 2019, McDonald et al., 2000; Rummukainen & Mendonça, 2016).

One could therefore hypothesize that, in real-world environments, where competing stimuli are presented continuously, with varying degrees of congruency, visual perception might be affected by auditory stimuli through the mediation of attentional factors and increased cognitive demand.

The two experiments reported here were intended to model domestic technologies such as television and gaming systems, which are currently able to present sound all-around the user, while keeping the visual stimuli limited to the field of view provided by the screen. It was our hypothesis that surrounding/immersive sound might sometimes have a detrimental effect to the visual experience, by redirecting attentional resources. In a previous study, where subjects were asked to focus on a visual event of the audiovisual reproduction, it was found that participants made more errors counting visual occurrences when the soundtrack was fully immersive and surrounding, as opposed to in less spatialized mono and stereo conditions (Mendonça et al., 2015). Participants observed larger areas of the screen and were therefore less able to focus specifically on the relevant screen area when more enveloping 3D sound was used. In the experiments reported here, we were specifically concerned about cognitive load and perceptual accuracy. Experiment 1 used realistic, immersive stimuli and manipulated the spatial distribution of the congruent sounds. Experiment 2 wished to replicate the effect found in Experiment 1, but in a more controlled setting. Random sequences of light flashes and noise bursts were presented, and participants had to focus on the visual events while a variety of sound distributions were presented. The cognitive load was assessed in both experiments through the indirect measure of pupil diameters.

VI. Experiment 1

2.1 METHODS

2.1.1 Participants

52 participants (11 females, 41 males) took part in this experiment. The participants had ages ranging between 20-50 years (31.68 +/- 6.84 years). All participants had normal vision or corrected-to- normal vision. All participants reported having normal hearing and were naive regarding the purpose of the study. All participants signed an informed consent to their participation. The experimental procedure and the collection of personal data were approved by the Aalto University Ethics Committee.

2.1.2 Settings

The tests were performed in an immersive audiovisual environment. The setting was inside an acoustically treated space and comprised three HD video projectors and 40 loudspeakers (Genelec 1029). The screens were 2.5 m by 1.88 m each and they were disposed in a pentagonal shape, where the two back sides of the pentagon were absent. The resolution of the video material was 4320x1080 px. A maximum of 12 loudspeakers we used for this test. Participants interacted with the system via a tablet with touch screen interface. This interface was developed specifically for the experiment with a Max/MSP visual programming language.

2.1.3 Stimuli

The visual stimulus consisted of a 9-second-long scene from a 3D movie provided by author Synes Elischka. The stimulus can be downloaded in Mendonça and Korshunova 2020. The movie was shot on 360 RED Epic cameras with a 6K high-quality footage. The sound was recorded with boom microphones and standard audio recording techniques for movie production. After the first second of the scene, there were falling boxes on the right side of the middle display. This stimulus was accompanied by four different audio conditions: mono, stereo, 5.1 and 7.1.4. All conditions were professionally mixed and mastered in a sound studio. The mono condition reproduced the soundtrack using only one loudspeaker, behind the main screen in the center. The stereo condition reproduced the soundtrack with two loudspeakers, one between screen A and B and the other between screen B and C (see Figure 2). The 5.1 and 7.4.1 conditions followed the standard loudspeaker distributions for these configurations. Condition 5.1 used 5 loudspeakers at the head level, one behind the screen, and a subwoofer. Condition 7.4.1 used 7 loudspeakers at head level, 4 above the head, and a subwoofer.

2.1.4 Procedures

Four groups of 13 participants performed the task, while watching the audiovisual scene with one out of the four audio conditions mentioned above. The participants were asked to count how many falling boxes they could see in the scene. During the task, the participant's gaze direction and pupil dilation were tracked by an eye-tracking system (Tobii Pro 2 Glasses).

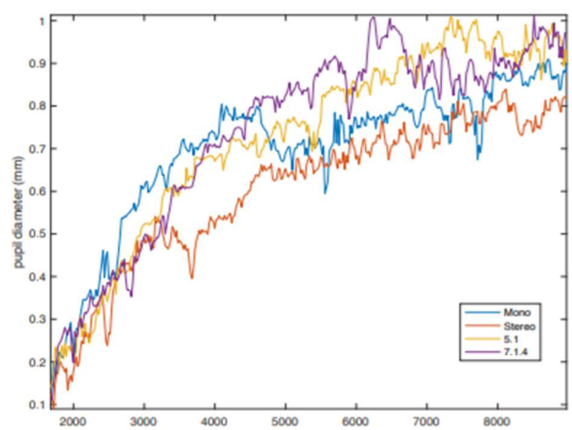
Each participant was seated at the center of the audiovisual system and had a tablet computer with a touch screen (Apple iPad Mini) in their hands. The participants were asked to perform a visual selective attention task: counting how many falling boxes they could see in the movie scene. The total number of falling boxes was 17. Before the start of the experiment, each participant practiced a trial with no audio and completely different visual stimuli which had the same average luminance. This allowed the participants' pupil diameter to stabilize.

2.2 RESULTS

In order to estimate the diameter of both pupils during the course of the presentation in each sound condition, both eyes were averaged per subject and missing frames filled with an interpolation function. To normalize pupil dilation across groups, the first second of the scene (before the counting task begun) was taken as baseline. Figure 1 shows how the average pupil diameter changed throughout time since the first second of the scene. The graph shows only the data from the moment boxes started falling onward. There was no change in screen luminance throughout the scene. As can be observed, pupils dilated in the first three seconds of boxes falling, then showing a trend toward stabilization. After one second of the onset of the task, some differentiation between conditions started to occur. Five seconds into the scene, when the diameter of pupils begun to stabilize, the lowest dilations occurred in the stereo condition, followed by the mono condition. Conditions 5.1 and 7.4.1 alternated in which led to more pupil dilation, overlapping curves at the end of the scene.

A one-way analysis of co-variance (ANCOVA) was conducted to determine statistically significant differences between the four sound conditions on pupil dilations, controlling for time. There was a significant effect of sound condition on pupil dilation after controlling for time ($F_{(3,12)} = 577.88$, $p = 0.000$). This test shows that the differences in progression throughout time per group are significant. Post-hoc pairwise comparison tests with a Scheffé correction show that the mono condition is different from all other conditions ($p \leq 0.001$). The stereo condition is different from mono and from 5.1 ($p \leq 0.001$) and from 7.1.4 ($p \leq 0.05$). In 5.1 results are different from all other conditions ($p \leq 0.001$). In 7.1.4, results are different from mono and 5.1 at ($p \leq 0.001$) and from the stereo ($p \leq 0.05$). Based on these findings, it is possible to conclude that there was more cognitive load in the 5.1 and 7.1.4 audio conditions. The stereo condition was cognitively the easiest sound condition.

Figure 1. Difference in pupil diameter throughout time, per sound condition.



VII. EXPERIMENT 2

3.1 METHODS

3.1.1 Participants

17 subjects (6 females, 11 males), with ages comprised between 21 and 44 took part in the experiment. None reported any known hearing impairment and all had normal or corrected-to-normal vision.

All experimental procedures were approved by the Aalto University Ethics Committee. All participants signed a written informed consent for their participation and data usage.

3.1.2 Setup and Stimuli

The experiment was run in a small anechoic chamber equipped with a multichannel spherical loudspeaker system. There were three LED yellow lights positioned at 0 deg in elevation, one at the center - 0 deg azimuth - and the remaining at +/- 20 deg in azimuth. The subject sat at the center of the room, 3 m away from the lights and loudspeakers, and black thin acoustically transparent curtains covered the sight of the loudspeakers and lights. The lights were sized 0.42 deg of the visual field and their luminance was 41 cd/m^2 at the viewing point. The experiment took place in a dark environment. The visual stimuli consisted of sequences of single 60 ms light flashes and the sound stimuli consisted of sequences of 60 ms white noise bursts. During the sequences of light flashes, each light flashed randomly in either of the three possible positions: left, right, or center.

Each trial presented one sequence of flashes and bursts. The number of flashes per sequence varied randomly, with a minimum of 5 and maximum of 13. The time between each visual flash was 150 ms. Noise bursts were not co-localized with the light flashes. There were 5 sound conditions. In h10, noise bursts occurred randomly either at -10 deg, 0 or +10 deg azimuth, always fixed at 0 deg elevation, and were always synchronized with the visual flashes; In h20, the noise bursts could occur additionally at -/+ 20 deg, the condition was otherwise similar to condition h10; In the h45 condition they occurred randomly at -45, -30, -20, 0, +20, +30, or +45 deg; Condition h90 was similar to h45, but sounds could additionally occur at -/+ 60 and -/+ 90 deg (11 loudspeakers); Condition h180 was similar to h90, but sounds could also occur at - 120 ,-150, 180, +150, and +120 deg.

3.1.3 Procedure

The sequences always ended with a light flash in either the left or right side, followed by a final flash in the center. Subjects were instructed to pay attention to the light and to report in which side the light has flashed last before the end of the sequence. They were told that they should ignore the sounds and report only where the light was perceived. Responses were reported on a keyboard. Pressing the left arrow indicated that the last light event was perceived to the left, and the pressing right arrow meant it was perceived in the right side. The answering time allowed was 1 sec. There were a total of 20 repetitions per experimental condition. All conditions were presented randomly within the experiment. Subjects were allowed to take breaks. On average, participants took 30 minutes to complete the experiment.

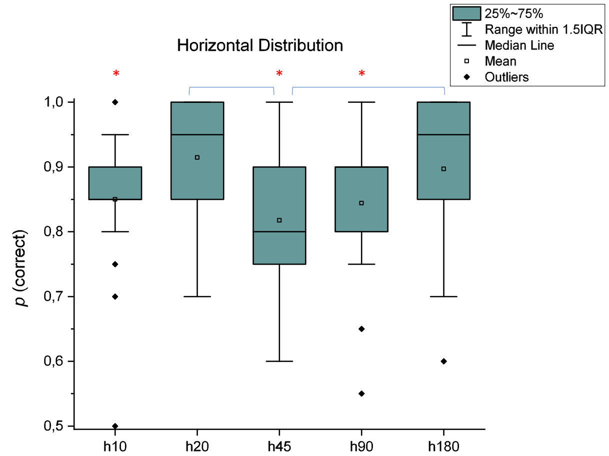


Figure 2. *Proportion of correct answers per sound condition.*

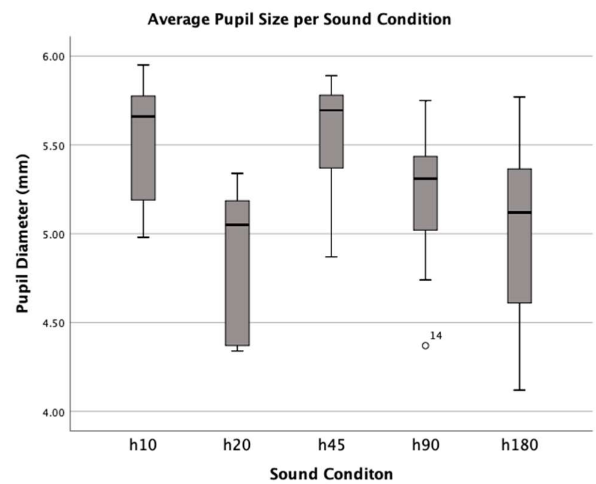


Figure 3. *Average pupil size per sound condition.*

3.2 RESULTS

Results regarding response accuracy are presented in Figure 2. Accuracy was highest in condition h20, the sound condition where the noise bursts could occur in the same region as the visual stimuli (Mean = 0.91, SEM = 0.02). Response accuracy was the lowest when sounds occurred within a distribution close to, but not the same as, where the light events were. Indeed, there were poorer lateralization scores in the h10 (Mean = 0.85, SEM = 0.03), h45 (Mean = 0.82, SEM = 0.03) and in the h90 (Mean = 0.84, SEM = 0.03) conditions. The portion of correct results was higher in the h180 condition, where sound events occurred all around the subject (Mean = 0.90, SEM = 0.03). The statistical tests confirmed significant differences between conditions ($F_{(2.77)} = 4.30$, $p = 0.011$). The post-hoc tests revealed that the h45 condition was significantly different from the h20 condition ($p = 0.006$). The h45 condition was significantly different from the h180 condition ($p = 0.03$).

Figure 3 shows the average pupil size in each of the sound conditions. The pupil size results confirmed the response accuracy results. The condition where the smallest pupil diameter occurred was h20 (Mean = 5.18, SEM = 0.12). The experimental conditions where the largest pupil size occurred were h10 (Mean = 5.32, SEM = 0.20) and h45 (Mean = 5.45, SEM = 0.31). A One-Way ANOVA revealed that there was a significant effect of sound condition on pupil diameter ($F_{(4)} = 15.85$, $p = 0.011$).

Post-hoc Tukey tests revealed that the results of the h20 condition were significantly different from the h10 and h45 conditions, but not from the remaining conditions. This might be due to the large variability of pupil sizes at the h90 and h180 conditions. These results indicate that sound distributions matching the area where visual stimuli can occur lead to lower cognitive load than other distributions; conversely cognitive load is significantly larger when sound distributions are close to, but not the same as, the area of visual events.

VIII. Discussion

In the experiments reported here, we were interested in quantifying the cognitive load during visual tasks accompanied by sounds with different spatial distributions.

In Experiment 1, stimuli were realistic and immersive, and sounds were always congruent. There were four sound conditions, where the enveloping sounds could occur in a narrow area (mono), or in increasingly wider areas (stereo, 5.1, 7.4.1). The task was to count the number of boxes falling in a given region of the screen. It was found that the conditions mono and stereo, where the sound occurred only within the area where visual stimuli were presented, led to significantly lower cognitive effort levels that the wider, more enveloping sound conditions.

In Experiment 2, the stimuli used remained complex in terms of number of events and audiovisual matching but were simplified otherwise. Visual stimuli consisted of sequences of light flashes and sounds were noise bursts. Lights occurred only at three positions (left, center, right), and the task was to indicate if the last flash of the sequence had been presented to the left or to the right. Sounds were presented synchronously, but always at random positions in space. The distribution of sound events in space varied across conditions, from a very narrow area (+/- 10 deg) to all around the participant. It was found that the best light flash localization accuracy was obtained when sounds occurred within the same region as the visual flashes. The worst accuracy was not obtained when sounds occurred all around the subject, but instead they were obtained when sounds occurred in distributions close-to, but not matching, the area of the visual flashes. Pupil dilations were significantly larger for those close regions, when compared to pupil dilations obtained in the condition with the same spatial distribution as the visual events. These results might signify that the distracting effect of sounds is greater when they occur just outside the attended area, as opposed to far away.

These results have direct implications for developers of audiovisual systems, such as gaming and television interfaces. Spatial sound presented outside of the area of visual interest might increase cognitive load and can potentially affect the perceptual experience.

Acknowledgements

This research was supported by the Center for Psychology at the University of Porto, Portuguese Science Foundation (FCT UIDB/00050/2020)

References

Driver, J., & Spence, C. (1998). Crossmodal attention. *Current opinion in neurobiology*, *8*(2), 245-253.

McDonald, J. J., Teder-SaÈlejaÈrvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature*, *407*(6806), 906-908.

Mendonça, C. (2020). Psychophysical models of sound localisation with audiovisual interactions. In *The Technology of Binaural Understanding* (pp. 289-314). Springer, Cham.

Mendonça, C., Rummukainen, O., & Pulkki, V. (2015). 3D sound can have a negative impact on the perception of visual content in audiovisual reproductions. *International Conference on Auditory Display,* (pp 138-143). Georgia Institute of Technology, http://hdl.handle.net/1853/54119

Catarina Mendonça and Victoria Korshunova. 2020. Published data from: Surround Sound Spreads Visual Attention and Increases Cognitive Effort in Immersive Media Reproductions. https://doi.org/10.5281/zenodo.3945057

Rummukainen, O., & Mendonça, C. (2016). Task-relevant spatialized auditory cues enhance attention orientation and peripheral target detection in natural scenes. *Journal of Eye Movement Research*, *9*(1).

Störmer, V. S. (2019). Orienting spatial attention to sounds enhances visual processing. *Current opinion in psychology*, *29*, 193-198.

Spatiotemporal Brain Mechanism of Auditory and Tactile Time Shrinking

Takako Mitsudo[a,b]*, Naruhito Hironaga[a], Hiroshige Takeichi[c], Yoshitaka Nakajima[d], Shozo Tobimatsu[a,e]
[a] Department of Clinical Neurophysiology, Neurological Institute, Faculty of Medicine, Graduate School of Medical Sciences, Kyushu University
[b] Department of Neuropsychiatry, Graduate School of Medical Sciences, Kyushu University
[c] Advanced Data Science Project (ADSP), RIKEN Information R&D and Strategy Headquarters (R-IH), RIKEN
[d] Department of Human Science/Research Center for Applied Perceptual Science, Faculty of Design, Kyushu University
[e] Department of Orthoptics, Faculty of Medicine, Fukuoka International University of Health and Welfare
mtaka@med.kyushu-u.ac.jp

Abstract— Complex brain networks are involved in the formation of an internal cognitive "clock" that we sought to explore by monitoring brain activity. Related to the "time-shrinking" illusion that results from any of the temporal assimilations, specifically, the duration of two successive intervals (T1 and T2) marked by three stimuli can be perceived as equal even when they are not. We have reported the spatiotemporal profiles of the human brain via auditory temporal assimilation using magnetoencephalography (MEG), focusing on encoding the temporal-parietal junction (TPJ), and judging the inferior frontal cortex (IFG; Hironaga, et al., 2017). However, the dependence property of sensory input is not clarified exclusively using one type of modality, i.e., auditory. We investigated the spatiotemporal profile of temporal assimilation using vibrotactile inputs. Along with the previous study, we measured human brain activities using MEG and tactile stimuli with temporal patterns of T1 and T2 of 160/240, 240/240 and 320/240. Behavioral results showed almost identical tendency as auditory time shrinking. Specifically, assimilation highly occurred when T1 was 80 ms shorter than T2. A distributed source-localization analysis was performed using minimum-norm estimates. Regional activity differences between judgment and no-judgment conditions were compared by addressing somatosensory and judgment-related regions, such as SI, SII, IFG, TPJ, motor, and premotor areas. The results showed that the right IFG and right TPJ were activated during tactile temporal judgment, and the activation profiles of these areas were quite similar to those of the auditory modality. These two regions are modality independent and may be related to time-perception (illusion) processing. Moreover, the right premotor area was activated specifically in response to tactile stimulation. Our results potentially represent cues to clarify the common and specific aspects of time perception, including time-shrinking illusions.

Keywords— Time-shrinking illusion | somatosensory | decision making | magnetoencephalography | premotor cortex

Temporal and Frequency Resolution Needed for Auditory Communication: Comparison between Young and Senior Listeners Utilizing Mosaic Speech

Yoshitaka Nakajima[1,2,3], Tatsuro Onaka[2], Akinori Oyama[3], Kazuo Ueda[1,2] and Gerard B. Remijn[1,2]

[1] Department of Human Science, Kyushu University, Fukuoka, Japan
[2] Department of Acoustic Design, Kyushu University, Fukuoka, Japan
[3] Sound Corporation, Fukuoka, Japan
yoshitaka.nakajima@100years.life

Abstract— A speech intelligibility experiment was performed to obtain more insight into the temporal and frequency resolution needed for Japanese speech communication. We used so-called "mosaic speech" developed in our research group (Nakajima et al., 2018; Santi et al., 2020), in an experimental paradigm that can reflect the spectro-temporal resolution necessary for listeners to perceive speech. In a previous study, in which only the temporal resolution of mosaic speech was varied, it was revealed that temporal resolution of about 40 ms is needed to convey linguistic information of Japanese speech. In the present study, 1) both temporal and frequency resolution were varied, 2) temporal resolution was varied in finer steps than before, and 3) young students and senior citizens were employed for comparison.

Japanese noun phrases of 7-13 morae were used as speech materials, and they were mosaicized in some different conditions: Sound-energy distribution on the time-frequency plane was divided into small blocks of given temporal and frequency block widths (i.e., temporal and frequency resolution) and the signal portion separated into each block was replaced with a noise portion of the same amount of energy in the same block area. Temporal resolution was varied in steps of 10 ms between 20 and 80 ms, and frequency resolution was varied as 1, 2, and 3 critical bandwidths. Given that our primary purpose was to estimate intelligibility conditions expressed as mora percent correct of 75% for young listeners with normal hearing, only a subset of all possible conditions were employed.

For young students (21-25 years of age; N = 15), intelligibility, defined as mora percent correct, always monotonically decreased as a function of the temporal block width if frequency block width was fixed. Intelligibility also monotonically decreased as a function of the frequency block width if the temporal block width was fixed. The 75% point was at temporal block widths of 51, 42, and 22 ms if the frequency block width was 1, 2, and 3 critical bandwidths, respectively. The results for the temporal resolution obtained at a frequency block of 1 critical bandwidth were similar to our previous results, but it was a new finding that the intelligibility decreased gradually when the temporal block width was increased in steps of 10 ms. A trade-off between temporal and frequency resolution was observed in the present stimulus conditions: Increasing the frequency block width by 1 critical bandwidth was equivalent to increasing the temporal block width by ~15 ms.

Intelligibility for senior citizens (65-79 years of age; N = 10), was systematically lower but basically the same trade-off was observed. Roughly speaking, the results for senior citizens were as if the temporal block width expanded by 15 ms, or as if the frequency block width expanded by 1 critical bandwidth. Thus, the present paradigm indicated the aging of the auditory system quantitatively, but differently from conventional audiometry: as a systematic decline of temporal and frequency resolution.

References

Nakajima, Y., Matsuda, M., Ueda, K., & Remijn, G. B. (2018). Temporal resolution needed for auditory communication: Measurement with mosaic speech. *Frontiers in Human Neuroscience*, 12, 149, doi:10.3389/fnhum.2018.00149.

Santi, Nakajima, Y., Ueda, K., Remijn, G. B. (2020). Intelligibility of English Mosaic Speech: Comparison between Native and Non-Native Speakers of English. *Applied Sciences*, 10, 6920. doi:10.3390/app10196920.

The Effect of Stimulus Duration on the Perceived Congruency of Stimuli Consisting of a Gabor Patch and an AM- or FM-Tone

Natalia Postnova[1] and Gerard B. Remijn[2]
[1]Graduate School of Design, Kyushu University, 4-9-1 Shiobaru, Minami-ku, Fukuoka 815-8540, Japan
[2]Department of Human Science, Faculty of Design/Research Center for Applied Perceptual Science, Kyushu University, 4-9-1 Shiobaru, Minami-ku, Fukuoka 815-8540, Japan
nataliapostnova@gmail.com; remijn@design.kyushu-u.ac.jp

Abstract— Research in the area of audiovisual perception often utilizes stimuli consisting of a Gabor patch and a modulated tone. Previous studies demonstrated cross-modal correspondence between the spatial frequency of the patch and the carrier or modulation frequency of the tone (Evans and Treisman, 2010; Green et al., 2019; Guzman-Martinez et al., 2012; Orchard-Mills et al., 2013a). Our recent study, however, showed that in the presence of dynamic changes in the Gabor patch, such as patch-flicker or grating drift, varying the spatial frequency of the patch or the carrier frequency of the tone had no or little effect on perceived congruency (Postnova et al., 2020). The sound modulation type, by contrast, did affect perceived congruency. In particular, in stimuli where the flickering frequency of the Gabor patch and the modulation frequency of the tone were similar, stimuli containing an AM-tone were perceived as more congruent than those with an FM-tone. This was demonstrated, however, only in stimuli in one duration condition (2 s), while most previous studies had used shorter or longer stimuli.

In the current experiment we therefore investigated the perceived (in)congruency in patch-tone stimuli of various durations. Gabor patches with two spatial frequencies (2 and 4 cycles per degree) in flickering (2 or 4 flickers per second) and drifting (0.5 and 2.0 degrees per second) modes were combined with an AM- or FM-tone with a 1000-Hz carrier frequency and 2- or 4-Hz modulation frequency. These audiovisual stimuli were 1, 2 or 4 seconds long. Participants rated perceived (in)congruency of each stimulus using a 7-point Likert scale (from 1 – incongruent to 7 - congruent). Average perceived (in)congruency ratings were analyzed separately for the stimuli with flickering Gabor patches and for the stimuli with drifting Gabor patches.

The results demonstrated that stimuli with a flickering Gabor patch and an AM-tone were rated as more congruent than stimuli with an FM-tone, regardless of the duration of the stimuli. Besides, stimuli with a flickering Gabor patch demonstrated similar rating patterns for stimuli of all durations. The analysis of the stimuli with drifting Gabor patches, however, suggested that the amount of temporal information might need to reach a certain minimum for a given parameter to modulate perceived (in)congruency. Stimuli consisting of a drifting Gabor patch and a 2-Hz modulated tone were given higher (in)congruency ratings as stimulus duration increased. Furthermore, the stimuli with a 4-Hz modulated tone demonstrated a clearer rating pattern, i.e., stimuli with a similar temporal frequency of the patch (temporal frequency = speed x spatial frequency) and the modulation frequency of the tone were more congruous than stimuli with dissimilar frequencies. However, also in stimuli with a drifting Gabor patch no clear effect of duration on perceived congruency was found. For any stimulus duration, the dynamic parameters in the stimuli, i.e., the temporal or flickering frequency of the patch and the modulation frequency of the tone had a strong influence on perceived congruency.

Keywords— Audiovisual integration, congruency, Gabor patch, tone modulation

References

Evans, K. K. and Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features, *Journal of Vision*. 10, 6. DOI:10.1167/10.1.6.

Green, J. J., Pierce, A. M. and Mac Adams, S. L. (2019). Multisensory integration is modulated by auditory sound frequency and visual spatial frequency, *Multisensory Research*. 32, 589–611. DOI:10.1163/22134808-20191402.

Guzman-Martinez, E., Ortega, L., Grabowecky, M., Mossbridge, J. and Suzuki, S. (2012). Interactive coding of visual spatial frequency and auditory amplitude-modulation rate, *Current Biology*. 22, 383–388. DOI:10.1016/j.cub.2012.01.004.

Orchard-Mills, E., van der Burg, E. and Alais, D. (2013). Amplitude-modulated auditory stimuli influence selection of visual spatial frequencies, *Journal of Vision*. 13, 6. DOI:10.1167/13.3.6.

Postnova, N., Nakajima, Y., Ueda, K., & Remijn, G. B. (2020). Perceived Congruency in Audiovisual Stimuli Consisting of Gabor Patches and AM and FM Tones, *Multisensory Research*, 34(5), 455-475. DOI: https://doi.org/10.1163/22134808-bja10041

Feathers, Metal and Weight Expectations

Helen E. Ross
Psychology, University of Stirling FK9 4LA, Scotland
h.e.ross@stir.ac.uk

Abstract— Sighted observers lifted a small metal weight, a bar of soap and a bag of loosely packed feathers, and estimated their heaviness relative to a hidden weight. The true weights ranged from 109-118 g (a barely discriminable difference). The metal was judged heaviest, the soap lighter and the feathers lightest. This order conforms to the size-weight illusion, not the material-weight illusion. All objects felt lighter when hefted on the hand than when lifted by a string. The results do not support the theory that natural objects should have their expected weight and be non-illusory.

Keywords— Size-weight illusion, material-weight illusion, weight expectations, feathers, density

I INTRODUCTION

Some familiar weight illusions are the size-weight illusion (SWI) (Charpentier 1891) and the material-weight illusion (MWI) (Wolfe 1898; Seashore 1899). In the SWI the larger of two objects of the same weight and surface material but different volume feels lighter when lifted. In the MWI, when two objects of equal weight and volume but different surface material are lifted, the apparently denser object feels lighter. The MWI is a much smaller illusion than the SWI (Buckingham & Goodale, 2013). Size contributes more than material to the perceived weight of objects when lifted (Vicovaro & Burigana, 2017). These illusions are often described as due to contrast with the expected weight.

Many aspects of weight perception are reviewed by Jones (1986) and weight illusions are reviewed in more detail by Buckingham (2014), Dijker (2014) and Saccone and Chouinard (2019). Buckingham classifies the explanations as follows. *Sensorimotor errors*: Larger objects are lifted with greater force, causing a mismatch between efference and afference. However, observers rapidly adapt their lifting force to an appropriate level, and the illusion persists. *Bottom-up effects:* Observers perceive density but report it as weight. It is not clear how density is perceived. Differences in rotational inertia may affect apparent weight for some types of objects, but not for the typical hand-held SWI. *Top-down effects*: The SWI and MWI result from contrast with the expected weight. There may be long-term prior expectations, or short-term conditioning. The illusions are 'cognitively impenetrable' (Pylyshyn, 1984) – they occur even when the observer knows that the objects have the same weight. Buckingham concludes that top-down expectations account for about half the strength of the SWI, the remainder being due to bottom-up effects such as grip force, lifting the object in the hand, or jiggling it to sense torques.

If we perceive density rather than weight, how is this done? Thouless (1931, p.17) suggested that apparent weight was a compromise between real weight (indicated by muscular and tactual stimuli) and density (given by a combination of cues to weight and visual indications of size). Thus, the cues to density are complex, and 'density constancy' is much less effective than other types of perceptual constancy. He suggested a formula for density-constancy which gave a ratio of about 0.3, much smaller than the value of about 0.7 found for size-constancy. Huang (1945) suggested that the ratio should be subtracted from unity, and the remainder would represent weight-constancy rather than density-constancy, giving a value of about 0.7. Izzet (1934) found that judgements of weight when the volume was changed were more accurate than judgements of density, showing that apparent weight is nearer to physical weight than it is to density.

It is sometimes asked "Which feels heavier, a pound of lead or a pound of feathers?" If the objects are packed in a normal manner the feathers will occupy a much larger volume than the lead, and this might cause a SWI for a sighted observer: the feathers should feel lighter. However, if the materials are visible, the feathers would be expected to be lighter and should feel heavier when lifted according to the MWI. Given that the SWI is much stronger than the MWI, the combined effect should favour the SWI. Scripture (1897, p.282) maintained that size dominated for a sighted observer, and that lead felt heavier than feathers: "Try it with a pillow and a piece of lead pipe. No matter if the scales do say that they weigh just the same, the pound of lead is much the heavier as long as you look at it."

No experiments have been reported in which observers lifted materials that were normally packed and visible. Wagman *et al.* (2007) used two cardboard boxes of equal size, one stuffed tightly with goose feathers and the other partially filled with lead shot in a plastic bag taped to the bottom of the box. The boxes weighed 637.9 g. The observers were blindfolded. The experimenter placed each box in turn on the palm of the observer's hand, in a balanced order. The observer hefted them and compared them several times and decided which felt heavier. The results showed that the lead box felt slightly heavier. The mass distribution of the lead box was bottom-heavy, and the authors concluded that the asymmetric mass distribution made it feel heavier (Amazeen & Turvey, 1996).

This tells us nothing about the role of expectations for sighted observers and natural objects. I argued (Ross, 1969) that if objects had their expected weight they should not be illusory, and should feel equal to a hidden object of the same weight. In that experiment weighted tins and polystyrene blocks of various sizes were compared with hidden weights. The objects were suspended inside a box and were lifted by vertical handles by the thumb and forefinger. Half the box was open to allow the test objects to be seen, and half was screened to cover the hidden weight. The polystyrene blocks ranged from 90-165 g and matched the hidden standard of 121 g when their density was about 0.14. A light set of tins ranged from 60-180 g and matched a hidden standard of 120 g when their density was about 1.39. A larger set of tins ranged from 160-270 g and matched a hidden standard of 210 g when their density was about 2.02.

These results might be taken to imply that sighted observers should be able to judge the true weight of a range of familiar objects. Babington-Smith (1965, p.149) investigated this and found fairly good rank ordering of weight for a variety of natural objects, but there were exceptions for objects of unexpectedly high or low density (e.g. a shotgun cartridge and a papier-mâché eggcup). High-density objects were judged too heavy, and low-density objects too light. This might mean that the observers' expectations were incorrect, or that expectancy theories are inadequate.

I therefore conducted an experiment (under Covid restrictions) using feathers and other objects available to me at home.

II. EXPERIMENT

Participants

The participants were 12 friends or neighbours of the author, 8 men and 4 women, aged between 55 and 81 years. All were right-handed. They were recruited and tested while outdoors for social gatherings.

Apparatus

There were three visible test objects, and a hidden weight, with weights ranging from 109-118 g. A photograph of the objects is shown in Figure 1. The test objects, in order of size, were:

1. A small brass weight from a domestic weighing scale set, marked as 4 oz (113 g). It was a cylinder with an area of 9.621 cm2 and a height of 1.4 cm. The volume of 13.47 cc gives a density of 8.39 g/cc. A string was tied round its centre and secured with transparent sellotape.
2. A bar of Imperial Leather hand soap, weighing 118 g. It was approximately cuboid in shape, measuring about 8 X 5 cm2 with a height of 2 cm. The volume of 80 cc gives a density of 1.48 g/cc. The soap was kept in its transparent plastic covering. String was tied around its centre and secured with transparent sellotape.
3. A large transparent plastic bag of down feathers weighing 109 g. The bag was approximately cuboid in shape, measuring about 37 X 27 cm2 with a height of 9 cm. The volume of 8991 cc gives a density of 0.01 g/cc. String was tied around its centre and secured with transparent sellotape.

Figure 1. The test objects from left to right: Jar with hidden weight, feathers, soap, brass weight.

There was also a hidden weight: a plastic bag of granulated sugar (110 g), lifted by a string inside an opaque cylindrical jar. The jar was dark blue (Hornsea Pottery). It had an area of 132.73 cm2 and a height of 15 cm, giving a volume of 1991 cc. The bag of sugar was approximately cuboid, measuring about 8 X 5 X 3 cc. The volume of 120 cc gives a density of 0.92 g/cc.

Procedure

The test objects were placed on a table. The participants stood in front of the table to lift the objects. They were told that the objects were fairly similar in weight, and the purpose of the experiment was to see how the method of lifting affected their apparent heaviness.

There were two methods of lifting: String or Hand. For the String condition the participants lifted all the objects by the string just above the point of attachment, using the thumb and forefinger of their right hand, and moved it gently up and down a few centimetres.. For the Hand condition they picked up the object with their left hand and placed it on their right hand, then hefted it gently.

The hidden weight was always lifted by the string, raising it a few centimetres within the jar. The participants were told that the hidden weight had a value of 10, and that they were to compare all the visible objects with it and give them a numerical value. They should say 10 if the object felt the same, 11 or 9 if it felt slightly heavier or lighter, 20 if it felt twice as heavy and 5 if it felt half as heavy. They lifted the hidden weight first and then the test object, but were allowed to repeat the procedure once or twice if uncertain of their judgement.

Half the participants were tested with the String condition first and half with the Hand condition first. Each participant was given one of the possible 6 orders of lifting the objects, using the same order in both conditions.

III. RESULTS AND DISCUSSION

Numerical estimates of heaviness, compared to a hidden weight of 10, are listed below for the 12 participants.

String				Hand		
	Metal	Soap	Feathers	Metal	Soap	Feathers
1	11	8	4	9	6	2
2	11	7	3	11	2	2
3	15	11	5	13	8	3
4	15	8	5	15	12	7
5	15	9	4	14	8	2
6	14	12	4	14	8	2
7	12	3	2	12	3	3
8	9	4	3	8	2	2
9	11	8	3	12	9	2
10	11	10	7	10	8	2
11	20	8	7	14	8	5
12	11	10	7	6	4	3
Mean	12.9	8.2	4.5	11.5	6.5	2.9
Median	11	8	4	11.5	8	2

When lifted by a string, the brass weight felt heavier than the hidden weight, the soap lighter and the feathers very much lighter. When hefted on the hand these effects were present, but the objects felt lighter than when lifted by a string. The data were not suitable for parametric statistics, so Wilcoxon's non-parametric test was used. Metal was judged significantly lighter when on the hand than when on a string ($W(8) = 5$, $p < .05$, 2 tails), as was soap ($W(11) = 5$, $p < .01$, 2 tails) and feathers ($W(12) = 3$, $p < .01$, 2 tails).

The reason why objects felt lighter on the hand than when lifted by a string was probably because the pressure per unit area was lighter when the object was spread over the hand. Weber (1834) noted that "a given point is pressed with less force by a weight that presses over a larger area……[and it] arouses the same sensation as metals with a lower specific weight." (Ross & Murray, 1996, p.63). Similarly Abbott (1864, p. 37) and Charpentier (1891) stated that an object feels lighter when it covers a larger contact area. Flanagan and Bandomir (2000) also found that objects feel lighter when lifted with a larger grip or a larger contact area.

The SWI clearly occurred for both methods of lifting. The prediction that normal visible objects should match a hidden weight was not upheld. The metal weight was judged slightly heavier and the soap slightly lighter, and the feathers very much lighter. The discrepancies between the estimates for the various objects and the hidden weight might be partly due to the discrepancies in true weight between all the objects. However, the Weber fraction for lifted weights at 100 g is about 10% (Ross & Brodie, 1987). The physical discrepancies were all less than this, so they were unlikely to affect the numerical estimates. If the objects had been ranked for true weight the soap should have been heaviest, followed by metal, hidden weight, then feathers. The observed order was metal, hidden weight, soap, feathers.

It could also be argued that the hidden weight did not act like a true hidden weight behind a screen, because the jar containing it was clearly visible. If the observers imagined that they were lifting the jar rather than an object inside the jar, this would affect the apparent weight of the hidden object. The volume of the jar lay between that of the soap and feathers, so on the basis of the SWI the apparent heaviness of the hidden weight should also lie between them. This was clearly not the case.

It can be asked whether the strength of the illusion differed with the method of lifting. It might be predicted that the illusion should be greater when the objects are held in the hand, because more tactile information is available about the size of the object. The strength of the illusion can be measured by the difference between the heaviness estimates for metal and feathers. For the means, the difference was 8.4 for String and 8.6 for Hand. For the medians it was 7 for String and 9.5 for Hand. This suggests that there was a slightly larger illusion when lifted on the hand, but the effect was not statistically significant ($W(9) = 9$, $p > .05$, 1 tail). A larger number of participants might confirm the results of Ellis and Lederman (1993) who found that the haptic SWI was stronger than the visual SWI.

In conclusion, for sighted observers lifting naturally packed objects, feathers do feel lighter than metal even though both possess approximately equal weights. This finding is consistent with the SWI, not the MWI. It is inconsistent with the idea that observers should have correct expectancies for natural objects and suffer no weight illusions.

Acknowledgements

I am grateful to my friends and neighbours who took part in the experiment, and to Professor David Murray for helpful comments on a draft of this paper.

References

Abbott, T.K. (1864*) Sight and Touch*. London: Longman, Green, Longman, Roberts, and Green.

Amazeen, E.L., Turvey, M.T. (1996). Weight perception and the haptic size-weight illusion are functions of the inertia tensor. *Journal of Experimental Psychology: Human Perception and Performance* 22: 213-232.

Babington-Smith, B. (1965) *Laboratory Experience in Psychology*. Oxford: Pergamon Press.

Buckingham, G. (2014). Getting a grip on heaviness perception: a review of weight illusions and their probable causes. *Experimental Brain Research* 232: 1623-1629.

Buckingham, G., Goodale, M.A. (2013). Size matters: A single representation underlies our perceptions of heaviness in the size-weight illusion. *PLoS ONE*, 8, e54709.

Charpentier, A. (1891). Analyse expérimentale de quelques éléments de la sensation de poids. *Archive de Physiologie Normales et Pathologique*s 3: 122-135. [Translated in Nicolas, S., Ross, H.E., Murray, D.J. Charpentier's papers of 1886 and 1891 on weight perception and the size-weight illusion. *Perceptual and Motor Skills* (2012) 115: 120-141.]

Dijker, A.J.M. (2014) The role of expectancies in the size-weight illusion: A review of theoretical and empirical arguments and a new explanation. *Psychonomic Bulletin and Review* 21: 1404-1414.

Ellis, R.R., Lederman, S.J. (1993) The role of haptic versus visual volume cues in the size-weight illusion. *Perception and Psychophysics* 53: 315-324.

Flanagan, J.R., Bandomir, C.A. (2000) Coming to grips with weight perception: effects of grasp configuration on perceived heaviness. *Perception and Psychophysics* 62: 1204-1219.

Huang, I. (1945) The size-weight illusion in relation to the perceptual constancies. *Journal of General Psychology* 33: 43-63.

Izzet, T. (1934) Untersuchungen ueber Wahrnehmungsgegenstände. VI. *Archiv für die Gesamte Psychologie* 91: 305-318.

Jones, L. A. (1986) Perception of force and weight: theory and research. *Psychological Bulletin* 100: 29-42.

Pylyshyn, Z.W. (1984) *Computation and Cognition: Toward a Foundation for Cognitive Science* (p. 114, pp. 130-145). Cambridge, MA: MIT Press.

Ross, H.E. (1969) When is a weight not illusory? *Quarterly Journal of Experimental Psychology* 21: 346-355.

Ross, H.E., Brodie, E.E. (1987). Weber fractions for weight and mass as a function of stimulus intensity. *Quarterly Journal of Experimental Psychology* 39A, 77-88.

Ross, H.E., Murray, D.J. (1996) *E.H. Weber on the tactile senses (2nd edition)*. Hove: Erlbaum (UK) Taylor & Francis.

Saccone, E.J., Chouinard, P.A. (2019) The influence of size in weight illusions is unique relative to other object features. *Psychonomic Bulletin and Review* 26, 77-89.

Seashore, C.E. (1899). Some psychological statistics II. The material weight illusion. *University of Iowa Studies in Psychology* 2: 36-46.

Scripture, E.W. (1897). *The New Psychology*. London: Walter Scott Ltd.

Thouless, R.H. (1931) Phenomenal regression to the 'real' object. II. *British Journal of Psychology* 22: 1-30.

Vicovaro, M., Burigana, L. (2017) Contribution of surface material and size to the expected versus the perceived weight of objects. *Attention Perception and Psychophysics* 79: 306-319.

Wagman, J.B., Zimmerman, C., Sorric, C. (2007) "Which feels heavier – a pound of lead or a pound of feathers?"

A potential perceptual basis of a cognitive riddle. *Perception* 36, 1709-1711.

Weber, E.H. (1834) *De pulsu, resorptione, auditu et tactu.* Leipzig: Koehler.

Wolfe, H. (1898). Some effects of size on judgments of weight. *Psychological Revue* 5: 25-54.

Your Money or Your Life?!
Social (Psychological) Distance and Temporal Discounting

Jordan Richard Schoenherr
Department of Psychology, Concordia University
Department of Psychology / Institute of Data Science, Carleton University
JordanRichard.Schoenherr@concordia.ca

Abstract— Humans tend to discount future outcomes. In the present study, I examined the extent to which monetary and health outcomes were discounted for self, close other, and distant other. Results demonstrated that participants' discounting was affected by psychological distance in terms of both social distance and temporal distance for money. In sharp contrast, participants discounted health outcomes for self and other to a much lesser extent.

Keywords— psychological distance, temporal discounting, prospect theory, social psychophysics

I. Introduction

Psychophysics is dominated by the study of sensory and perceptual phenomena whether immediately available to the participants or in the form of remembered magnitudes (e.g., Algom, 1992; Magnussen & Greenlee, 1999). Recent decades have seen studies of comparatively psychophysical phenomena that are related to human interaction (i.e., social psychophysics; e.g., Jack & Scyns, 2017; Kurban, 2001). In an analogous manner to sensory and memory psychophysics, recent approaches to social psychophysics attempt to identify laws that relate information obtained from facial expressions to social perception, e.g., emotion. Despite the focus on sensory and perceptual stimuli, there is no issue (in principle) with extending social psychophysics into the domain of memory.

One means to do this is to consider interpersonal closeness (i.e., social distance) in terms of an abstraction representation of distance between two entities. A recent demonstration of effects of psychological distance was conducted by O'Connell et al. (2013). Using a temporal discounting paradigm, participants were provided with the option of accepting an immediate reward or deferring their reward for a larger amount in the future. By varying the recipient of the reward (self, close other, or distant other) they found that social distance and temporal distance of the reward both impacted judgments.

The purpose of the present study was to 1) replicate findings that both temporal and relational distance result in discounting using a modified paradigm from O'Connell et al., (2013) and 2) examine whether monetary outcomes were discounted at a higher level than health outcomes. Given that the study occurred during the SARS-CoV2 pandemic, this study also statistically controlled for the importance of health beliefs.

Prospect Theory and Temporal Discounting

Prospective Theory (PT) is one of the most influential theories in the social sciences for explaining widespread violations of expected utility (Kahneman & Tversky, 1979; Tversky and Kahneman, 1986). The theory is used to assess how participants value losses and gains for immediate and delayed rewards. It assumes that individuals are loss averse, leading to less discounting for potential losses than gains, and that high- and low-probability events are under- and overestimated, respectively. Subsequent observations that PT was limited to choices defined by a limited number of risky outcomes lead to modifications (e.g., Tversky & Kahneman, 1992). PT remains an influential theory outside of psychological sciences, e.g. behavioural economics (Barbeis, 2013; Simonsohn, 2014).

Despite concerns over assumptions of PT (e.g., Gal & Rucker, 2018), recent studies continue to support its general claims. Most notably, Ruggeri et al. (2020) conducted a large multi-national study across 19 countries and 13 languages and found that the general patterns of loss aversion were observed across countries while also noting that there were differences between individuals and societies. This suggests that while loss aversion is a generalizable phenomenon, individual differences and social norms and conventions determine the degree of loss aversion a participant's experiences.

Social Distance and Social Psychophysics

Memory psychophysics assumes that a systematic relationship can be established between the remembered physical magnitude of a stimuli and the subjective magnitude. Thus, following encoding into memory, a representation can be retrieved and compared to a representation that is immediate available in the environment. Humans also retain representations of other humans in terms of physical magnitudes (e.g., height and weight) as well as social magnitudes (e.g., status differences; A. P. Fiske, 1993; A. P. Fiske et al., 1991). For A. P. Fiske (1991), relational models that describe the properties of relationships are analogous to types of scales, i.e., Stevens' scale typology. Thus, in addition to representations of specific social distance based on relational closeness, individuals also retained schematic relationships between themselves and others. Social psychophysics can then assume that the perceived magnitude of a potential reward will be a function of two dimensions of psychological distance, namely, temporal and social.

A recent demonstration of this was provided by O'Connell et al. (2013). In their study, they assumed that outcome devaluation was a result of experiencing less empathy for the recipient (e.g., Lowenstein, 1996). In a single experiment, participants were required to make a choice between a small, immediate (or in the near future) reward and a comparatively larger delayed reward. They were told that the recipient of the reward would either be themselves, a close other, or a distant other. In order to prime other, participants were asked to type out the individuals name and this name was then presented on the screen in the close and distant other conditions. In this case, they computed the indifference point (Johnson & Bickel, 2002) for delays of weeks or months for rewards up to £100. Their analysis revealed the hypothesized kind of discounting: in addition to a nonlinear decreasing function describing the relationship between value and time, they also found that greater discounting occurred for a distant other relative to a close other and that both of these were discounted more than decisions made for themselves.

II. Present Study

The study conducted by O'Connell et al. (2013) leaves open a number of questions. Specifically, how are participants conceptualizing 'closeness'? Closeness could be considered in terms of similarity. Conversely, it would be considered in terms of familiarity. The present study explores these issues by modifying O'Connell et al. (2013) paradigm in a number of meaningful ways. First, rather than priming participants with the name of the close or distant other on a trial-by-trial basis, participants are asked to type out the name and then rate the perceived similarity and familiarity they have with the recipient. Second, participants are presented with a fixed set of rewards and time periods rather than […]. Finally, in addition to replicating the monetary reward, participants were also provided with rewards framed as health outcomes, i.e., $100 or 100% health. Here, I assume that monetary outcomes will be discounted to a greater extent than health outcomes.

III. Methods

Participants.

The present study examines pilot data from a small convenience sample (*n*=10) of undergraduate university students. Students received 0.25% toward their grade in an undergraduate psychology course for participants in the online study.

Materials.

Replicating O'Connell et al.'s original methods, the decision for 'self' took place before the decision for other. This fixed order was adopted to avoid self-bias. However, the order of blocks for close and distant other conditions was counterbalanced.

In contrast to the original study, participants were primed in advance of each block of trials to indicate their level of closeness with the person they listed. Participants were required to type the person's name and then indicated 1) how close they felt to the person (familiarity) and 2) how much contact they had with the person in terms of physical or virtual interaction (similarity). [Familiarity and similarity ratings were rated on a scale of X to Y.]

In contrast to the original study, the choice alternatives will be fixed in terms of both the time and the reward. On each trial, participants had to choose between $100 in 18-months from now or accept a lesser amount, i.e., $25 in 0 months (now), $50 in 6 months, or $75 in 12 months (see Figure 1). Participants were required to make a 'yes' or 'no' decisions.

Figure 1. Sample stimulus display presenting the possibility of a small, immediate reward ('Now') for monetary. Modified from stimuli developed by O'Connell et al., (2013).

A similar method was also used for health outcomes condition. On each trial, participants had to choose between 100% gain in health in 18-months from now or accept a lesser amount, i.e., 25% gain in health in 0 months (now), 50% gain in 6 months, or 75% gain in 12 months. Participants were required to make a 'yes' or 'no' decisions.

Closeness Manipulation. The closeness manipulation entailed presenting participants with a set of priming instructions followed by a series of ratings. Priming instructions consisted of three types:

Distant Block: "Now, we want to imagine someone that you do not have a close personal relationship with, i.e., an 'acquaintance'. This could be someone a co-worker, someone you see at a grocery store, or anyone that you do not know all that well. When you have a person in mind, please type their first name below. If you do not know their first name, you can type their role below.

We will not save this response. It is only being used to help you focus on that person."

Close Other Block: "Now, we want to imagine someone that you have a close personal relationship with. This could be close friend or familiar member. When you have a person in mind, please type their first name below. Please type their first name below.

We will not save this response. It is only being used to help you focus on [that person.]"

Self Block: "Now, we want to imagine making the decision for yourself. Please type your first name below.

We will not save this response. It is only being used to help you focus on [yourself.]"

Closeness and Similarity Ratings. Beneath each one of the priming instructions, a blank field allowed participants to type in the name of the other or self. They were then asked to rate their familiarity and similarity.

Procedure. In advance of the block of experiment, participants were provided with a general set of instructions:

"This task involves a series of choices. On some trials, you will make choices between smaller amounts of money now or larger amounts of money later. On other trials you will make between minor *increases in* health now or larger *increases in* health later. However, you will also be asked to perform the task as if you were someone you know. Try and put yourself in their shoes and imagine how they would respond.

When making choices, do not consider your real financial or health circumstances."

Participants will be presented with one of three sets of priming instructions concerning distant, close, or self. Whereas the order to distant and close other will be randomize, the order of the *'self'* block will be fixed. Each block *was* further subdivided into questions concerning finance and health. Health and finance questions will be randomized.

The primary task was followed by a secondary distractor task which required participants to rate social categories and complete individual difference questionnaires. Following this task, participants were then asked to rate their overall concerns with health, the coronavirus, and money. They were also asked to indicate whether they had received a vaccination and rate their level of employment.

Due to a change in the Qualtrics interface, a coding error was introduced. This resulted in the presentation of only 3 experimental blocks, leading to participants being presented with two of one block (e.g., monetary) and one of the other (e.g., health).

IV. RESULTS

Prior to analysis, participants' ratings of similarity and familiarity for 'Close' and 'Distant Other' were compared to make difference scores indicating the amount of distance between the two individuals they were using for this task.

The principal dependent variable for this analysis was the devaluation score. The devaluation score was computed by subtracting the outcome selected by the participants from the maximum outcome at 18 months, e.g., Devaluation = Highest Outcome – Selected Outcome. ANCOVA included Familiarity and Similarity Difference Scores as covariates along with the variables time, relation closeness, and outcome type (money or health). Individual difference measures were not included due to the small sample size (n = 10) and the variation associated with individual differences.

A notable finding of the present study was the relationship between similarity and familiarity scores and patterns of devaluation. First, a marginal interaction was observed between Choice Type, Temporal Interval, and Familiarity Difference Score, $F(2,14)$ = 3.05, MSE = 23.46, p = .082. This pattern qualifies the effect of Familiarity Difference Score, $F(1,7)$ = 17.6, MSE = 246.39, p = .004.

As Figure 2 illustrates, participants were least likely to select the devalued response alternative for health outcomes than they were for monetary outcomes such that the functions for health remained relative stable over the available temporal intervals. Moreover, patterns of devaluation were nearly identical for 'Close Other' and 'Self' and devaluation was significantly less than that for 'Distant Other'. This results partially replicates the finding obtained by O'Connell et al. (2013): the greater the social (psychological) distance between 'Self' and 'Other' the greater the corresponding level of devaluation.

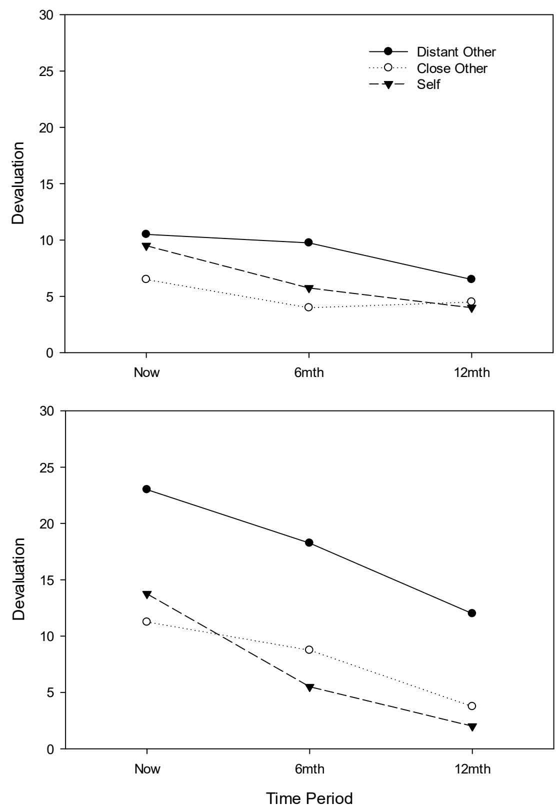


Figure 2. Discounting patterns for health (top) and money (bottom) for close other, distant other, and self.

A significant relationship was observed between Relationship Closeness and Similarity Difference Score, $F(2,14)$ = 4.26, MSE = 502.65, p = .039. Namely, discounting was reduced for those individuals who selected an 'Other' that was rated as highly similar to themselves. In social psychological terms, this likely reflects a closer 'in-group' member. This results also qualifies a marginal main effect of Relational Closeness, $F(2,14) = 2.86$, $MSE = 502.65$, $p = .095$.

Similarly, Time and Rated Familiarity Difference also interacted, $F(2,14) = 5.60$, $MSE = 113.47$, $p = .026$: greater levels of discounting were observed with larger differences in the familiarity ratings between close and distant other.

Finally, to qualify the relationship between familiar and devaluation, the overall devaluation score was computed for each participate and a correlation was obtained with the familiarity difference score. As Figure 3 demonstrates, the greater the difference in familiarity between close and distant other, the greater the level of devaluation

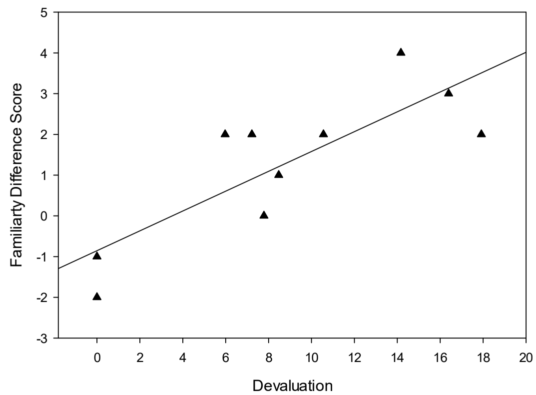


Figure 3. Relationship between difference in rated familiarity of 'Close' and 'Distant' Other and monetary and health devaluation.

V. CONCLUSIONS

The current study was directed toward replicating and expanding the results of O'Connell et al. (2013). In their study, participants produced patterns consistent with temporal discounting. Participants devalued outcomes to the extent that they were close to recipient of the reward such that they discounted outcomes that 'distant' others would receive more so than 'close' others, and the outcomes of close others more so than those that they would receive themselves. Despite some methodological differences (discussed below), the current study replicated the devaluation of outcomes from 'Distant' others. However, unlike O'Connell et al.'s findings, outcomes for 'Close' others remained comparable to those for the 'Self'.

The current study also extends these findings into the domain of health outcomes. First, health outcomes were not discounted to the same extent as monetary outcomes. Specifically, participants resisted accepting small, immediate improvements in health in a manner that differed from monetary outcomes. In contrast, there was more discounting in medium-term periods (12 months). Thus, participants would accept a 5% reduction in health outcomes if it meant obtaining these outcomes sooner. This might of course have been related to the pandemic, i.e., participants might have been keenly aware that obtaining the majority of possible outcomes sooner rather than later was more desirable.

Caveats and Future Directions.

First, the results of the present study must be viewed conservatively given that they reflect data from a small, pilot study. As a number of individual differences are likely important in determine a participant's willingness to delay reward (e.g., Mischel et a., 1988), the small sample size of the pilot might distort the relationships between variables. Second, the methodological differences between the current study and previous studies might also have led to different results. For instance, it is possible that the failure to replicate the result that there was greater devaluation for 'Close Other' and 'Self' observed by O'Connell et al., (2013) might have been a result of methodological differences of reminding participants on each trial about the 'Close' and 'Distant' other than had selected.

If methodological difference are a concern, it should be noted that the results should look quite different. Namely, there should be less of a difference in discounting pattern between 'Close' and 'Distant' Other, and 'Close' Other should be associated with a greater difference in discounting relative to 'Self'. The fact that the current study produced the opposite pattern suggests that the 'Close' Other that was selected is valued more than those selected in O'Connell et al.'s paradigm. Moreover, despite the likely benefit of trial-by-trial priming, the current study directly assessed two psychological dimensions (familiarity and similarity) and controlled for them statistically.

The results of the present study provide additional insight into Prospect Theory and studies of temporal discounting more generally. First, it might be the case that the utility associated with health is simply greater than that associated with money. On this account $100 is negligible relative to even small gains in health (e.g., 25% gains). This is compounded when we consider the real, and perceived, severity of the global public health crisis in which this study was conducted. Some evidence against this comes from a subsequent study (Schoenherr, in preparation) that increased the value from $100 to $1000. In this case, no differences were observed between the patterns of performance: the pattern of monetary and health discounting remained largely unchanged, although more discounting (5%) was observed in the $1000 study.

REFERENCES

Algom, D. (1992). Memory psychophysics: An examination ofits perceptual and cognitive prospects. In D. Algom (Ed.), *Psychophysical Approaches to Cognition* (pp. 441-513. Elsevier.

Barberis, N. C. (2013). Thirty years of Prospect Theory in economics: A review and assessment. Journal of Economic Perspectives, 27, 173-96.

Fiske, A. P. (1991). *Structures of social life: The four elementary forms of human relations: Communal sharing, authority ranking, equality matching, market pricing*. Free Press.

Fiske, A. P. (1993). Social errors in four cultures: Evidence about universal forms of social relations. *Journal of Cross-Cultural Psychology*, 24, 463-494.

Fiske, A. P., Haslam, N., & Fiske, S. T. (1991). Confusing one person with another: what errors reveal about the elementary forms of social relations. Journal of personality and social psychology, 60(5), 656.

Gal, D., & Rucker, D. D. (2018). The loss of loss aversion: Will it loom larger than its gain?. Journal of Consumer Psychology, 28(3), 497-516.

Jack, R. E., & Schyns, P. G. (2017). Toward a social psychophysics of face communication. Annual review of psychology, 68, 269-297.

Johnson, M. W., and Bickel, W. K. (2002). Within-subject comparison of real and hypothetical money rewards in delay discounting. *Journal of the Experimental Analysis of Behavior*, 77, 129–146.

Kurzban, R. (2001). The social psychophysics of cooperation: Nonverbal communication in a public goods game. Journal of Nonverbal Behavior, 25, 241-259.

Loewenstein, G. (1996). Out of control: visceral influences on behavior. Organ. Behav. Hum. Decis. Proc. 65, 272–292.

Magnussen, S., & Greenlee, M. W. (1999). The psychophysics of perceptual memory. *Psychological Research*, 62(2), 81-92.

Mischel W., Shoda Y., Peake P. K. (1988). The nature of adolescent competencies predicted by preschool delay of gratification. *Journal of Personality and Social Psychology*, 54, 687–696.

Petrusic, W. M., Baranski, J. V., & Kennedy, R. (1998). Similarity comparisons with remembered and perceived magnitudes: Memory psychophysics and fundamental measurement. *Memory & Cognition*, 26(5), 1041-1055.

Ruggeri, K., Alí, S., Berge, M. L., Bertoldo, G., Bjørndal, L. D., Cortijos-Bernabeu, A., ... & Folke, T. (2020). Replicating patterns of prospect theory for decision under risk. *Nature human behaviour*, 4(6), 622-633.

Simonsohn, U. (2014). '[15] Citing Prospect Theory'. Data Colada, February 10. Available at http://datacolada.org/15.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5, 297-323.

Modality Specificity of Amodal Completion

Hiroshige Takeichi[1], Keito Taniguchi[2] and Hiroaki Shigemasu[2]
[1]Open Systems Information Science Team, Advanced Data Science Project (ADSP), RIKEN Information R&D and Strategy Headquarters (R-IH), RIKEN
[2]School of Information, Kochi University of Technology
takeichi@a.riken.jp; shigemasu.hiroaki@kochi-tech.ac.jp

Abstract— The perception of occluded parts of objects is called amodal completion. It does not contain any visually-identified surface property in its output but is best predicted by the output namely *a posteriori* three-dimensional arrangement and solid shape, which is potentially improved by additional non-visual information e.g., touch. A virtual reality experiment was conducted with fifteen student participants. Word stimuli were presented in a head-mounted display for recognition but masked partly with a ribbon virtually positioned either in front of or behind the word so that they appear partly occluded or cut off, respectively. The visual depth cue was either provided by binocular stereopsis or limited by monocular presentation. The haptic cue was either missing, provided consecutively provided or concurrently, by actively tracing a real off-screen bar which was positioned correspondingly with the virtual ribbon. Recognition performance should be improved by amodal completion and was therefore compared between the depth cue conditions in a mixed linear model with a logistic link function. Random factors comprised participant, stimulus elevation which was randomized across trials, trial number to index perceptual learning and word and their interactions. As a result, word recognition was better with the stereoscopic cue that indicated occlusion, but not with the haptic cue, although both cues were effective in depth perception. It suggests that amodal completion is associated with high-level visual representation up to the three-dimensional stage exhaustively, but is confined exclusively to be modality-specific, in exact accordance with the nature of occlusion: three-dimensional and optical.

Keywords— Amodal Completion, Humans, Stereopsis, Haptic Perception, Depth Perception, Virtual Reality

Acknowledgment

This study is supported by Grant-in-Aid for Scientific Research (C) JP20K03500 from the Japan Society for the Promotion of Science.

On the Density of Faces in Face Space: Explanation in Terms of Infinite Dimensional Riemannian Geometry

James T. Townsend[1], Cheng-Ta Yang[2] & Hao-Lun Fu[2]
[1]Indiana University; [2]National Cheng Kung University

Abstract – Two important papers of the late 1990's and early 2000s by James Tanaka and colleagues (Tanaka, Giles, Kremen & Simon, 1998; Tanaka & Corneille, 2007) put forth the hypothesis that a repository of face memories can be viewed as a vector space. This hypothesis broadens the thesis of Tim Valentine that face space is constituted of feature vectors in a finite dimensional vector space (e.g., Valentine, 2001). Within that space, each face can be thought of as propagating an attractor field surrounding the point representing a specific face. Tanaka and associates do not emphasize actual dynamics but do hypothesize that the attractor field is 'larger' for less familiar faces than for familiar faces. The math interpretation would be that for a given distance from a less familiar face-point, the vectors for that field would be larger than that for a familiar face. So, if a point is picked midway between two candidate faces, observers are predicted to judge that midway face as being more similar to the less familiar stimulus point. Their experimental evidence supported this prediction. We propose an alternative interpretation based on our earlier propounded infinite dimensional Riemannian Face Manifold (Townsend, Solomon, Wenger & Spencer-Smith, 2001). That study contained an example potentially closely related to the Tanaka concerns, that of 'the other-race effect'. We feel this approach avoids some of the issues involved in the gradient theme by working directly with the type of metric inherently associated with the face space.

Smell Infant Psychophysics: Methodological adaptation for COVID-19 Pandemic

Rosana M Tristão[1], Lucas Lauand[1], Kassandra S F Costa[1], Luiz A Brant[1], Geraldo M Fernandes, Karina N Costa[1], Jan Spilski[2], & Thomas Lachmann[2,3]
[1]Faculty of Medicine and University Hospital, *University of Brasília, Brasília, Brazil*
[2]Center for Cognitive Science, University of Kaiserslautern, Germany
[3]Facultad de Lenguas y Educación, Universidad Nebrija, Spain
rmtt@unb.br and lachmann@rhrk.uni-kl.de

Abstract— Olfactory loss in adults is one of the initial and most frequent acute clinical manifestations of SARS-CoV-2 infection. The few studies that have evaluated this alteration in the paediatric range have shown that children have less olfactory sensory loss than adults. This study aims to assess the olfactory sensory perception of new-borns of women who previously tested positive for COVID-19 during pregnancy compared to new-borns of women who did not test positive for COVID-19 during pregnancy. It also aims to develop and validate a behavioural evaluation scale of olfactory sensory-perceptual perception in new-borns. This work is based on the scientific literature on the sensory-perceptual development of smell in foetuses and new-borns. New-borns will be exposed to four odours: breast milk, vanilla (sweet), coffee (bitter) and distilled water (neutral).

Keywords— new-born, perception, smell, COVID-19, SARS-CoV-2

I. INTRODUCTION

Pleasant or unpleasant responses to odours or tastes have been observed since the early stages of individual development (Schaal, 1988; Steiner, 1979) and even in premature babies (Goubet et al., 2002). Nasal and oral trigeminal subsystems are anatomically functional by the 4th gestation week, thus human nasal and oral chemoreceptors undergo anatomical development from very early ages. Informative odour can theoretically be transduced to the early brain, though perceptual processing may not be complete by the last gestational trimester (Bremner, Lewkowicz, & Spence, 2012). Unfortunately, testing of human fetuses is still difficult and findings may be solely interpreted as possible evidence for prenatal odour processing (Schaal et al., 2004).

Smell is an inevitable and ubiquitous stimulation source from the first steps of individual development, since nasal chemo sensors develop before other sensory systems (with the exception of somaesthetic / kinaesthetic sensors) and are in direct contact with stimuli that result from the biological functioning of a mother-baby relationship (Bremner et al., 2012). Thus, the foetus is able to process olfactory stimuli present in the mother's diet while still in the womb and then retain this information for months after birth suggesting odour memory and exhibit a preference for these odours (Mennella et al., 2003). In addition, it has been shown that the foetus is capable of associating stimuli from different sensory modalities, both internal and external (Smotherman & Robinson, 1990), and retain these associations for around three days after birth. There is also evidence for higher neonatal sensitivity for human odours, such as sweat from adults, and evidence that male newborns are more sensitive to different levels of dilution in these kinds of odorants (Loos et al., 2017). Several methodologies have been described to evaluate the newborns' olfactory perception based on behavioral responses (e.g.: Allam et al., 2006; Engen et al., 1963) and physiological responses (e.g.: Anunziata et al., 2020; Bartocci et al., 2000).

II. PRESENT STUDY

The goal of the present study is to develop and validate a behavioural evaluative scale of olfactory perception in healthy new-borns and to apply this scale to new-born children of women infected with COVID-19 during pregnancy comparing to new-born children of women without COVID-19 infection history. This is a retrospective comparative analytical cohort study of 300 new-borns exposed and unexposed to COVID-19 during pregnancy. The data collection will follow the experimental procedure in a previous study that explored odours of the maternal breastmilk, vanilla (sweet) and distilled water (neutral). A coffee smell was implemented as an addition to this previous study to include the acid/bitterness category to the categories of stimuli.

III. METHODOLOGY

This study would take place at the outpatient clinic for new-borns born to pregnant women with COVID -19 (and Regular Growth and Development Outpatient clinic with new-borns born to pregnant women without Covid-19). The data collection will follow an experimental procedure from Bartocci et al (2000 that explores odours of the maternal breastmilk, vanilla (sweet), coffee (acid/bitterness) and distilled water (neutral). The coffee smell is an addition to the previous study to include the acid/bitterness category to the categories of stimuli. Each test epoch consists of 30 s of baseline definition followed by 30 s of smell exposure with a two-minute interval for washout effect (Figure 1).

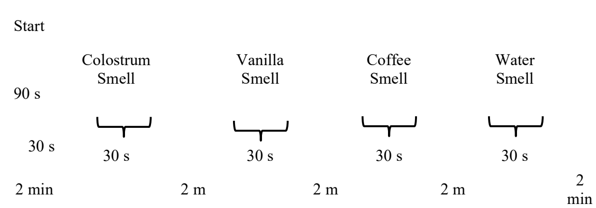


Figure 1. Timeline of odour exposure during one session adapted from Bartocci et al (2000).

The research assistants will wear surgical gloves for each handling of the solutions, and they will be prepared in an auxiliary room to avoid the spread of the smell of the solutions in the environment. Each solution will be kept in a hermetically sealed flask before and during all session to avoid smell impregnation. The room temperature will be maintained between 19-21ºC. The baby will be accommodated and cuddled on the lap of a familiar caregiver to prevent social stress and in a calm, relatively quiet (between 40 and 50dB(A)) environment reserved inside the Hospital Paediatric Follow-Up facility. The infants behavioral state during the session will be monitored and registered for each stage of smell exposure ranging from state 1 (deep sleep) to sate 6 (crying). In case of the baby achieving state level six (cry – intense cry or high motor activity) or in need of a diaper change, the session will be paused and resumed as soon the baby returns to the state five or less. Mothers will be instructed to not use moisturizers or perfumes and to bring the babies already breastfed at least 15 min before the session.

Image coding:

1. Phase I coding: will be carried out for the development of the olfactory sensory evaluation scale. The videos will be analyzed frame by frame to survey categories of behavioural responses to the four different olfactory stimuli. The rating levels will consider two response categories "yes" or "no", for present and absent reaction to the smell. Factor Analysis will be performed with analyses of the main components for the construction of the instrument and analysis of the sample data.

2. Phase II coding: will be carried out for comparative evaluation of new-born children of pregnant women with and without a diagnosis for Covid-19.

IV. Newborn infant Psychophysics: Clinical application

It is feasible to argue the hypothesis of the involvement of the foetus' olfactory bulb during intrauterine life as one of the indelible pathophysiological manifestations to the clinical diagnosis of COVID-19 with neurosensory olfactory deficit in foetuses and new-borns affected by intrauterine infection. This study aims to investigate if new-born children of women infected with COVID-19 during pregnancy have olfactory sensory changes. The clinical trial was registered in the Brazilian Registry of Clinical Trials (ReBEC- RBR-65qxs2).

References

Allam, M. D.-E., Marlier, L., & Schaal, B. (2006). Learning at the breast: Preference formation for an artificial scent and its attraction against the odor of maternal milk. *Infant Behavior and Development*, *29*(3), 308–321.

Anunziata, F., Fabiola Macchione, A., Sofia Mitrano, A., D'aloisio, G., Elena Ferreyra, M., David Pontoriero, R., … Carlos Molina, J. (2020). Respiratory and emotional reactivity to ethanol odor in human neonates is dependent upon maternal drinking patterns during pregnancy. *Drug and Alcohol Dependence*, *213*.

Bartocci, M., Winberg, J., Ruggiero, C., Bergqvist, L. L., Serra, G., & Lagercrantz, H. (2000). Activation of Olfactory Cortex in Newborn Infants After Odor Stimulation: A Functional Near-Infrared Spectroscopy Study. *Pediatric Research*, *48*(1), 18–23.

Bremner, A. J., Lewkowicz, D. J., & Spence, C. (Eds.). (2012). *Multisensory development* (1st ed). Oxford, U.K: Oxford University Press.

Engen, T., Kaye, H., & Lipsitt, L. P. (1963). Olfactory Responses and Adaptation in Human Neonate. *Journal of Comparative and Physiological Psychology*, *56*(1), 73—-.

Loos, H. M., Doucet, S., Védrines, F., Sharapa, C., Soussignan, R., Durand, K., … Schaal, B. (2017). Responses of Human Neonates to Highly Diluted Odorants from Sweat. *Journal of Chemical Ecology*, *43*(1), 106–117.

Mennella, J. A., Forestell, C. A., & Pepino, M. Y. (2003). The Flavor World of Infants. *Perspectives on Swallowing and Swallowing Disorders (Dysphagia)*, *12*(4), 10–20.

Schaal, B. (1988). Olfaction in infants and children: Developmental and functional perspectives. *Chemical Senses*, *13*(2), 145–190.

Schaal, B., Hummel, T., & Soussignan, R. (2004). Olfaction in the fetal and premature infant: Functional status and clinical implications. *Clinics in Perinatology*, *31*(2), 261— 85, vi—-vii.

Smotherman, W. P., & Robinson, S. R. (1990). The Prenatal Origins of Behavioral Organization. *Psychological Science*, *1*(2), 97–106.

Steiner, J. E. (1979). Human Facial Expressions in Response to Taste and Smell Stimulation. In *Advances in Child Development and Behavior* (Vol. 13, pp. 257–295).

Checkerboard Speech: A New Experimental Paradigm for Investigating Speech Perception

Kazuo Ueda[1,2], Riina Kawakami[2,3], and Hiroshige Takeichi[4]
[1]Unit of Perceptual Psychology, Department of Human Science/Research Center for Applied Perceptual Science/ Research and Development Center for Five-Sense Devices, Kyushu University
[2]Department of Acoustic Design, Faculty of Design, Kyushu University
[3]Present affiliation: Rion Co., Ltd.
[4]Open Systems Information Science Team, Advanced Data Science Project (ADSP), RIKEN Information R&D and Strategy Headquarters (R-IH), RIKEN
ueda@design.kyushu-u.ac.jp; riina0326@gmail.com; takeichi@riken.jp

Abstract— A new type of degraded speech stimuli, i.e., checkerboard speech (interrupted over time-by-frequency), was developed. The intelligibility of checkerboard speech and interrupted speech (interrupted over time), both of which retained a half of the original speech, was examined. The checkerboard speech stimuli with 20 frequency bands brought nearly 100% intelligibility irrespective of segment duration, whereas, with 2 and 4 frequency bands, a trough of 35%-40% appeared at the 160-ms segment duration. The intelligibility of interrupted speech stimuli decreased as segment duration increased. The results revealed the limitations of underlying auditory organization for speech cues scattered in a time-frequency domain.

Keywords— degraded speech, speech intelligibility, auditory organization

I. Introduction

A typical "cocktail party" situation requires a listener not only to segregate a stream of a target speech, but also to pick up and organize fragments of speech cues partly masked by the other sound (Best et al., 2019), into a coherent stream of the target speech. A similar situation can be created without employing other masking sound, but just by deleting some parts of the original speech [Fig. 1 (a) and (c)]. We named such degraded speech as *checkerboard speech* (Ueda, Kawakami, and Takeichi, 2021), after *checkerboard noise* invented by Howard-Jones and Rosen (1993).

Here we examine experimental results that compare the intelligibility of checkerboard speech and interrupted speech (e.g., Miller and Licklider, 1950) stimuli, providing some parts of the discussion (Ueda et al., 2021) and an additional note.

II. Methods

A total of 20 paid listeners participated in this study. Their normal hearing within the frequency range of 250-8000 Hz was ensured by the tests with an audiometer (Rion AA-56). The research was conducted with prior approval of the Ethics Committee of Kyushu University (approval ID: 70).

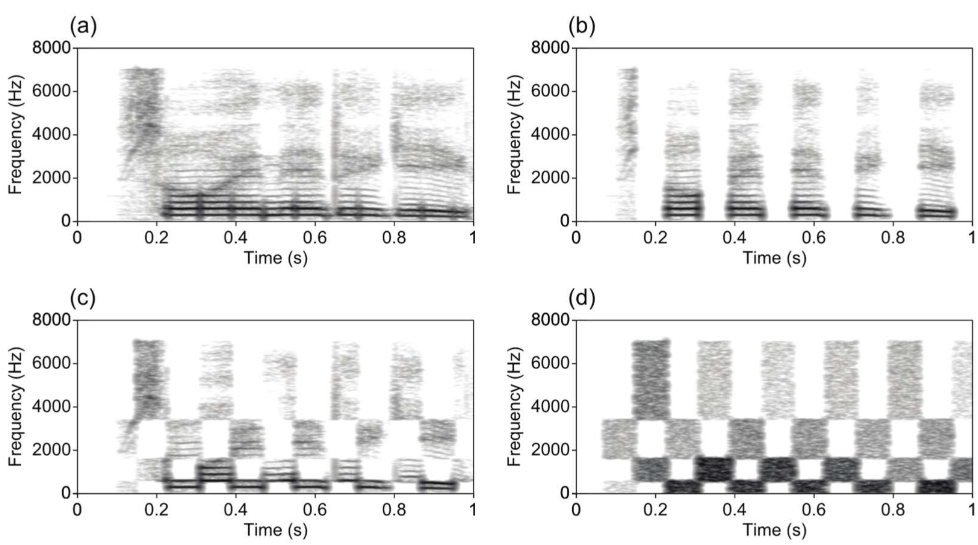


Fig. 1. Examples of narrowband spectrograms for degraded speech, produced from the same fragment of an original spoken sentence by a female talker. (a) Original speech, reconstructed with bandpass filtering into 4 frequency bands (passbands of 50-570, 570-1600, 1600-3400, 3400-7000 Hz) and segmenting with an 80-ms time window including 5-ms root-of-raised-cosine ramps in amplitude, (b) interrupted speech with the 80-ms segment duration, (c) checkerboard speech with the 4 frequency bands and 80-ms segment duration, and (d) checkered mosaic speech with the same time-frequency segmentation. From Ueda, Kawakami, and Takeichi (2021).

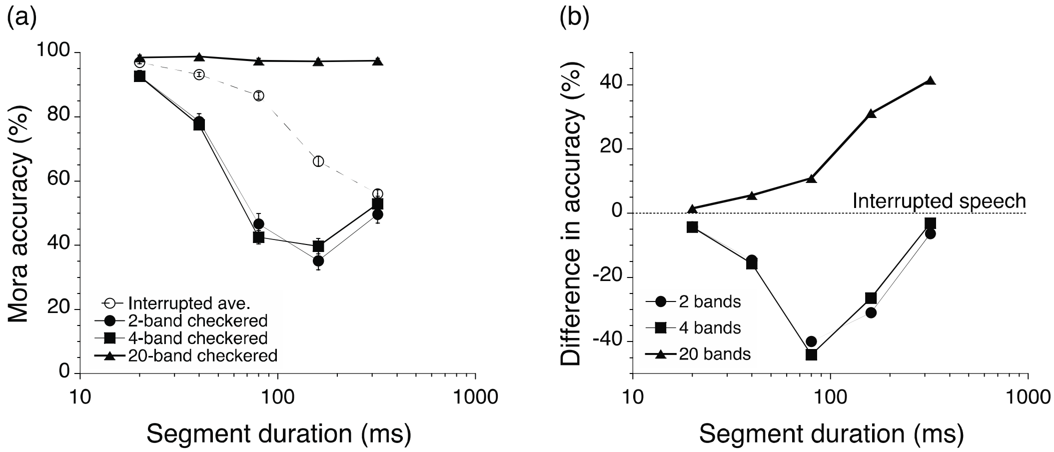


Fig. 2. The experimental results. (a) Mean percentages of mora accuracy (n = 20) for the checkerboard speech and interrupted speech stimuli as a function of segment duration and number of frequency band. The data for the interrupted speech stimuli were averaged over the number of frequency bands. (b) Mora accuracy differences between the checkerboard speech and interrupted speech stimuli. The percentages of correct for the checkerboard speech stimuli are represented as deviations from those for the interrupted speech stimuli at corresponding segment duration. Error bars in panel (a) reflect standard error of the mean (SEM). From Ueda et al. (2021).

A total of 150 Japanese sentences spoken by a female talker were extracted from the "Multilingual Speech Database 2002" (NTT Advanced Technology; 16000-Hz sampling, 16-bit linear quantization). The average duration per sentence was 2.5 s (SD = 0.39), and the average number of morae (a mora is a syllable-like unit in Japanese) per sentence was 18 (SD = 2.8). The extracted spoken sentences were edited to eliminate unnecessary blanks and noises. The edited speech samples were converted into 44100-Hz sampling, with 16-bit linear quantization.

The frequency range from 50 to 7000 Hz was divided into 2, 4, and 20 frequency bands with bandpass filter banks. The passbands of the filters for the two-frequency band condition were 50-1600, and 1600-7000 Hz, whereas for the four-frequency band condition, the passbands were 50-570, 570-1600, 1600-3400, and 3400-7000 Hz. The passbands for the 20-frequency band condition were determined according to the critical bandwidths (see, the filter bank B in Ueda and Nakajima, 2017). We adopted the four frequency bands determined by Ueda and Nakajima (2017).

Five steps of segment duration (20, 40, 80, 160, and 320 ms), three steps of frequency bands (2, 4, and 20), and two types of reduction (checkerboard and interruption) were combined to produce 30 conditions in total. A block of trials consisted of a set of full conditions. Five blocks of trials, in which the order of conditions was randomized in each block, were constructed. Sentences were randomly allotted to one of the trials for individual participants, who listened to the stimuli through headphones (Beyerdynamic DT 990 PRO) in a sound-attenuated booth. They were instructed to write down the morae that they could immediately recognize, and not to fill blanks afterwards from the context. Other procedural details were described in Ueda et al. (2021).

III. Results

The percentages of mora accuracy are represented in Figure 2(a). The results for the interrupted speech stimuli were averaged over the number of frequency bands. The intelligibility for the interrupted speech stimuli went down from almost perfect (97%) at the 20-ms segment duration to a little better than a half point (56%) at the 320-ms segment duration along with the segment duration, as expected from the previous investigations with meaningful speech sentences (e.g., Powers and Wilcox, 1977).

We additionally observed that the intelligibility for the 20-band checkerboard speech stimuli stayed at ceiling (97%-99%) irrespective of segment duration, whereas the intelligibility for the 2- and 4-band checkerboard speech stimuli was highest (93%) at the 20-ms segment duration, lowest (35% or 40%) at the 160-ms segment duration, and moderate (50% or 53%) at the 320-ms segment duration.

These observations were supported by the analysis using a generalized linear mixed model (GLMM) with a logistic linking function as implemented in an add-on for JMP (SAS Institute Inc., 2018). The data were analyzed for fixed effects of segment duration, number of frequency band, type of reduction (all categorical predictors), and their interactions, and for random effects of listener and sentence. This model revealed the p values smaller than 0.001 for a number of effects and interactions including segment duration [$F(4, 2860) = 128.19$], type of reduction [$F(1, 2798) = 12.78$], number of frequency band × type of reduction [$F(2, 2806) = 10.05$], segment duration × number of frequency band [$F(8, 2835) = 6.53$], and segment duration × type of reduction [$F(4, 2855) = 22.13$]. The p values were 0.020 in segment duration × number of frequency band × type of reduction [$F(8, 2839) = 2.29$] and 0.137 in number of frequency band [$F(2, 2802) = 1.99$].

Figure 2(b) shows the differences in the mora accuracy between the checkerboard speech and interrupted speech stimuli. The average accuracy for the interrupted speech stimuli at each segment duration is represented as the baseline, that is, "0" of the vertical axis. The deviations of the intelligibility for the 2- and 4-band checkerboard speech stimuli, reached the maximum at the 80-ms segment duration (-40% and -44%, respectively). The monotonic increase observed for the 20-band checkerboard speech stimuli essentially reflected the decrement of intelligibility for the interrupted speech stimuli.

IV. Discussion

Both the checkerboard speech and interrupted speech stimuli used in the present investigation retained a half of each original speech signal; nevertheless, in most of the cases, the intelligibility for these stimuli was not 50% [Fig. 2(a)]. The intelligibility for the checkerboard speech stimuli was affected by both the number of frequency bands and segment duration, with a wider range of variation (35%-93%) than the range for the interrupted speech stimuli (56%-97%). Furthermore, segment duration was the primary determinant for the intelligibility for the checkerboard speech stimuli with a small number (2 or 4) of frequency bands; on the other hand, the intelligibility for the stimuli with 20 bands was close to 100% irrespective of segment duration. Therefore, the results suggest that speech cue integration across critical bands occurred for the 20-band stimuli irrespective of segment duration, whereas the integration across frequency bands was difficult for the 2- and 4-band stimuli, especially at the 80- and 160-ms segment duration. Thus, the auditory system exhibits limitations in organizing spectrotemporally scattered speech cues, if the number of frequency bands is small. Furthermore, the limitations should be responsible for the intelligibility difference between the checkerboard speech stimuli with a small number of frequency bands and interrupted speech stimuli, which is particularly apparent at the 80-ms segment duration [Fig. 2(b)].

By contrast, for the interrupted speech stimuli, it has been well-known that intelligibility is high, 80% or above, at the segment duration shorter than 100 ms, and then declines gradually as the segment duration is extended, to about 50% at the segment duration close to or longer than the duration of a word, i.e., about 500 ms (Miller and Licklider, 1950). The present results are in line with this general trend.

Shafiro et al. (2018) showed that a minimum of word intelligibility appeared at the 250-ms segment duration for interrupted speech stimuli, when the stimuli were lowpass-filtered at 2000 Hz. They claimed that the segment duration at the dip could be estimated with a probability summation for how many words were sampled in each segment (mainly at low frequency of interruption) and how often a word was sampled (mainly at high frequency of interruption): At the minimum point, the proportion of words (i.e., words contained per each segment) is equal to the number of segments per word. The present authors introduced an analogous probability summation, taking a mora as a perceptual unit instead of a word, of which the average duration was about 140 ms (2,500 / 18). The model predicts the segment duration at the minimum intelligibility to be about 100 ms, which points to the bottom of the U-shaped curves observed for the checkerboard speech stimuli with a small number of frequency bands [Fig. 2(a)]. The successful prediction implies that, for spectrally degraded (coarsely segmented) speech stimuli, the segment duration at the minimum intelligibility may be determined by the average duration of the perceptual unit. The conjecture needs further verification in future investigations.

The variance in intelligibility observed for the checkerboard speech stimuli were very small, even in the most severely degraded conditions, i.e., at the 80- or 160-ms segment duration [Fig. 2(a)]. This is to be noted, because other degrading methods like local time-reversal (e.g., Ueda et al., 2017; Ueda and Ciocca, 2021), or desynchronizing channel slits (Silipo et al., 1999), often introduce much larger variance in intelligibility among participants in such a low intelligibility range. Further investigations are warranted to elucidate the cause of the difference.

Acknowledgments

This work was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI Grant No. JP19H00630. The authors would like to thank Yoshitaka Nakajima for providing J language software routines, Hikaru Eguchi for programming in LiveCode, Yoshitaka Nakajima, Gerard B. Remijn, Valter Ciocca, and Wolfgang Ellermeier for valuable discussion, and Josef Schlittenlacher and Valeriy Shafiro for providing insightful suggestions to improve the manuscript for the paper published.

References

Best, V., Roverud, E., Baltzell, L., Rennies, J., & Lavandier, M. (2019). The importance of a broad bandwidth for understanding "glimpsed" speech. *The Journal of the Acoustical Society of America*, 146(5), 3215-3221

Howard-Jones, P. A., & Rosen, S. (1993). Uncomodulated glimpsing in "checkerboard" noise. *The Journal of the Acoustical Society of America*, 93(5), 2915-2922.

Miller, G. A., & Licklider, J. C. R. (1950). The intelligibility of interrupted speech. *The Journal of the Acoustical Society of America*, 22(2), 167-173.

Powers, G. L., & Wilcox, J. C. (1977). Intelligibility of temporally interrupted speech with and without intervening noise. *The Journal of the Acoustical Society of America*, 61(1), 195-199.

SAS Institute Inc. (2018). JMP, version 14.3.0 [computer program].

Shafiro, V., Fogerty, D., Smith, K., & Sheft, S. (2018). Perceptual organization of interrupted speech and text. *Journal of Speech, Language, and Hearing Research*, 61(10), 2578-2588.

Silipo, R., Greenberg, S., & Arai, T. (1999). Temporal constraints on speech intelligibility as deduced from exceedingly sparse spectral representations. *Proceedings of ESCA Eurospeech99*, 2687-2690.

Ueda, K., & Ciocca, V. (2021). Phonemic restoration of interrupted locally time-reversed speech: Effects of segment duration and noise levels. *Attention, Perception, & Psychophysics*, 83(5), 1928-1934.

Ueda, K., Kawakami, R., & Takeichi, H. (2021). Checkerboard speech vs interrupted speech: Effects of spectrotemporal segmentation on intelligibility. *JASA Express Letters*, 1(7), 075204.

Ueda, K., & Nakajima, Y. (2017). An acoustic key to eight languages/dialects: Factor analyses of critical-band-filtered speech. *Scientific Reports*, 7(42468), 1-4.

Ueda, K., Nakajima, Y., Ellermeier, W., & Kattner, F. (2017). Intelligibility of locally time-reversed speech: A multilingual comparison. *Scientific Reports*, 7(1782), 1-8.

Helmholtz and the Impact of the Stereoscope on Visual Science in Germany

Nicholas J. Wade
Psychology Department, University of Dundee, Scotland
n.j.wade@dundee.ac.uk

Abstract— Helmholtz was born 200 years ago but his influence is still felt in visual science largely through his *Treatise on Physiological Optics.* It was translated into English to mark the centenary of his birth. The study of vision was transformed by Wheatstone's invention of the stereoscope and the instrument was eagerly adopted and adapted by Helmholtz. Stereoscopic depth perception posed problems for theories of binocular single vision developed by Müller, based on stimulation of corresponding retinal points. This pillar of German visual science was undermined by stereoscopic phenomena and by Wheatstone's empiricist interpretation of them. As was the case for his theory of perception, Helmholtz followed Wheatstone rather than his compatriots.

Keywords— physiological optics, Helmholtz's *Handbuch*, binocular single vision, stereopsis, rivalry

I. Introduction

Hermann Ludwig Ferdinand von Helmholtz (1821-1894) was born on August 31st 200 years ago. He was a towering scientist and represented the rise and ascendancy of German science. Instruments and institutions are named after him and he received many awards from universities and scientific societies. He was ennobled to von Helmholtz in 1883. The span of his research can be appreciated from his *Popular Lectures on Scientific Subjects* (Helmholtz, 1873, 1881). The topics surveyed included the conservation of force, the origins of the planetary system, axioms of geometry, Goethe's science and the organization of universities. Two series of lectures were on vision, one on its then recent advances and the other on its relation to painting. In the three parts of the *Handbuch der physiologischen Optik* (Helmholtz, 1867) vision was examined progressively with regard to the physics of the stimulus, the physiology of the sense organs, and the psychology of perception; these made up the three volumes of the *Treatise* (see Wade, 2021b). The title pages of the *Handbuch* and *Treatise* are shown in Figure 1, together with portraits of Helmholtz and Southall, who translated it into English. Here the focus is on Helmholtz's analysis of binocular vision and the impact that Wheatstone's observations had on visual science in Germany. As was the case for his epistemology, Helmholtz followed Wheatstone's interpretation of stereoscopic vision rather than those of his compatriots.

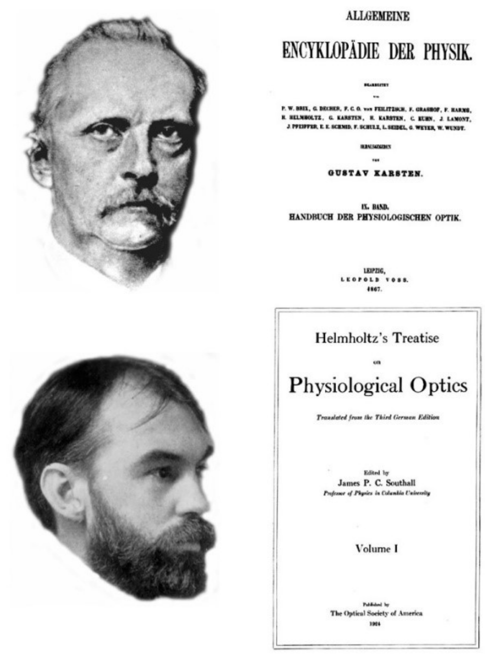


Figure 1. Upper left, detail of the portrait of Helmholtz from the frontispiece of Southall's translation, and lower left, detail of a portrait of James Powell Cocke Southall (1871-1962). Upper right, the title page of the first complete edition of Helmholtz's *Handbuch*, together with (lower right) the title page of Southall's translation.

II. Single vision with two eyes

Wheatstone (1838) wrote: "No question relating to vision has been so much debated as the cause of the single appearance of objects seen by both eyes" (p. 387). Binocular single vision has been discussed at least since the time of Aristotle and it has been examined experimentally after Ptolemy defined lines of visual correspondence for the two eyes (see Howard & Wade, 1996). Many of the early statements about binocular single vision are reflections of its breakdown and the experience of binocular double vision (Howard & Rogers, 1995; Wade, 1998). Helmholtz's mentor, Johannes Müller (1838), considered that he had supplied the solution with a geometrical analysis supported by detailed experiments: points on a circle passing through the rotation centres of each eye and the point of fixation stimulate identical retinal points and are seen as single. Vieth (1818) made a similar analysis earlier and it is now referred to as the Vieth-Müller circle. Thus, there were two states of binocular vision – singleness when points lie on the binocular circle and double vision otherwise. It is an irony of science that in the year Müller provided his final account of this circle of correspondence, Wheatstone announced his invention of the stereoscope and his experiments with it. Indeed, Wheatstone introduced the binocular circle to English-speaking scientists only to reject its validity: “objects whose pictures do not fall on corresponding points of the two retinæ may still appear single” (Wheatstone, 1838, p. 384). Not only were the objects seen as single but they were also in depth.

A summary of Wheatstone’s article was soon translated into German (Wheatstone, 1839), and a full translation was provided later (Wheatstone, 1842). William James (1890) shrewdly observed that Wheatstone’s paper “contains the germ of almost all the methods applied since to the study of optical perception. It seems a pity that England, leading off so brilliantly the modern epoch of this study, should so quickly have dropped out of the field. Almost all subsequent progress has been made in Germany, Holland, and, *longo intervallo*, America” (pp. 226-227). The impact of Wheatstone’s stereoscopic investigations on German visual science was seismic. Müller’s (1838) elegant analysis of binocular single vision was challenged and many of his countrymen sprang to its defense. Volkmann (1859a) summarized the sentiment: “If Wheatstone were right on this point, the doctrine of vision was threatened to be overturned completely” (p. 2). What did Wheatstone demonstrate that stirred visual scientists, particularly those in Germany (see Turner, 1994), so dramatically? First, he showed that points that are not on the Vieth-Müller circle can be seen as single and in depth; secondly, he presented a figure that stimulated identical retinal points and was seen as double. Helmholtz (1867) addressed both issues in part 3 of his *Handbuch*. Wheatstone’s observations touched another nerve in the context of German visual science: his descriptions of stereoscopic depth and its characteristics were not based on detailed quantitative experiments as was the case for evidence supporting the Vieth-Müller circle. Again, this was articulated by Volkmann: “One only needs to remember Johannes Müller’s excellent investigations into the identical retinal points and their positions in order to see that with the abandonment of the doctrine of identity, the doctrine of double images, of the horopter and the apparent distance between two images in the fields of vision collapse, precisely those doctrines which are based on exact experiments, namely experiments that can be controlled by measurements and numbers. It is therefore very natural for the physiologists to counter Wheatstone’s conclusions and try to reconcile the phenomena he observed with the theory of identical retinal sites” (1859a, p. 2). Helmholtz was in the mold of Wheatstone rather than of Müller and Volkmann.

III. Binocular vision

Helmholtz commenced his account of binocular vision by examining visual direction: monocular vision could signal direction alone but location required distance which could be supplied by binocular vision. In this context he added to the terminology by introducing the term ‘cyclopean eye’ (Wade, 2021a). This topic had been examined by Wells (1792) in what was perhaps the first book dedicated to binocular vision. Wells formulated principles of visual direction with two eyes that were rediscovered by Hering decades later (see Ono, 1981). Helmholtz emphasized that stereoscopic depth perception is learned, and that the invention of the stereoscope “made the difficulties and imperfections of the Innate Theory of sight much more obvious than before” (Helmholtz, 1873, p. 274). Thus, the stereoscope helped to give Helmholtz precisely what he needed to strengthen and defend the empirical theory of space perception against attacks on it by Hering (see Lenoir, 1994; Turner, 1994). He maintained that not only is binocular depth perception learned, but that all spatial perception is founded on judgmental acts based on experience.

Wheatstone had both mirror and prism stereoscopes made for him in 1832, publishing his account of the mirror stereoscope in 1838 (see Wade, 1983). Brewster (1849) made a more compact stereoscope using half-lenses which acted as both refractors and magnifiers so that paired stereoscopic pictures or photographs could be combined (see Wade, 2019). The instruments and their inventors are shown in Figure 2. Investigations of binocular vision were transformed by the invention of the stereoscope. It could be argued that the stereoscope heralded the revolution in vision and the instrument was embraced by Helmholtz. He initiated research on binocular vision in the 1850s although his experiments on binocular vision were undertaken in the early 1860s when in Heidelberg. In prosecuting his experimental enquiries Helmholtz modified Brewster’s lenticular stereoscope so that the separation of the lenses could be adjusted for different interocular separations; he also adapted Wheatstone’s mirror stereoscope so that disparities could be enhanced. This was achieved by extending the separations between the mirrors and he called the instrument a telestereoscope. The stereoscope was very important to Helmholtz, both for the experimental world it exposed and also for his inferential theory of vision.

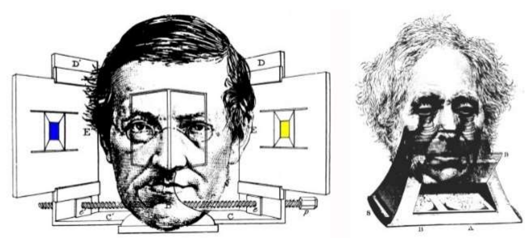


Figure 2. *Stereoscopists* by Nicholas Wade. Left, Wheatstone and his mirror stereoscope; right, Brewster and his lenticular stereoscope.

The first stereoscopes were based on mirrors, prisms, or lenses. The use of colours for separating the eyes to see depth was realised by a contemporary of Helmholtz (Rollmann, 1853), although they were not called anaglyphs until later in the century (Wade, 2021a). They were mentioned in the *Handbuch*: “He [Rollmann] draws two projections on the same black card, one with red lines, the other with blue. Then he takes a red glass in front of one eye and a blue glass in front of the other and only sees the red lines with that eye, with this only the blue ones, which can then be combined to form a relief. (Helmholtz, 1867, p. 685). That is, the red lines are seen as black through the blue glass as are the blue lines through the red glass. This text was incorrectly translated by Southall who wrote “he can see only the red lines through the red glass and the blue lines through the blue glass” (Helmholtz, 1925, p. 356). Anaglyphs are displays in which the left and right eye images are printed in different colours, such as red and cyan, and they are viewed through filters of the same colours. They have typically been used to present slightly different images to each eye so that they are seen in stereoscopic depth, but they can also be enlisted to demonstrate binocular rivalry and a range of other dichoptic interactions.

The stereoscope, perhaps more than other instruments invented in the 1830s, ushered in a transformation of studies of vision. The methods of physics could be applied to experiments on binocular vision: stimuli could be manipulated systematically and the ensuing responses recorded. Helmholtz placed greater emphasis on the former than the latter: stimuli took precedence over responses. Despite the fact he was familiar with Fechner’s (1860) book, Helmholtz remained with physics rather than embracing psychophysics. The most detailed early experimental studies of stereoscopic vision were conducted by Panum (1858) and Volkmann (1859a) and Helmholtz drew upon these extensively: Volkmann receives more citations in the *Handbuch* than either Wheatstone or Fechner. Volkmann was a colleague of Fechner at Leipzig and introduced him to investigations of the senses. In an extensive article, Fechner (1861) examined aspects of binocular vision, particularly binocular rivalry, and emphasised the role of attention. It is in this article that ‘Fechner’s paradox’ is described: a light appears brighter with one eye than with two when a neutral density filter is placed in front of one of them.

Brewster’s (1844) interpretations of stereoscopic depth were similar to those proposed earlier by Brücke (1841), a close associate of Helmholtz. Eye movements were implicated in order to reconcile the binocular phenomena with Müller’s theory of identical retinal points. If the eyes changed convergence rapidly while viewing solid objects then this could be the basis for perceived depth rather than the combination of slightly disparate retinal points in the two eyes. While Helmholtz adopted eye movement interpretations of many visual effects, this did not apply to stereoscopic vision. Dove (1841) illuminated stereoscopic pairs with brief electrical sparks which generated afterimages and depth was visible with them. Volkmann (1859b) presented paired stimuli briefly with a tachistoscope and this also resulted in stereoscopic depth; he invented the instrument specifically to xamine this question. Helmholtz confirmed these observations and wrote: “Both these experiments … show that ocular movements are not necessary for perception of depth; because after-images move with every movement of the eye, and it is simply impossible to make disparate images correspond to each other by any such movement” (1925, p. 456).

Binocular rivalry can involve competition between colours or contours and both were investigated before the invention of the stereoscope (see Wade & Ngo, 2013). The combination of different colours presented to corresponding regions of each retina became an issue of theoretical importance following experiments on colour mixing: are colours combined by either eye as they are when selected from the spectrum? Helmholtz, like Wheatstone, embraced colour rivalry as evidence in favour of an inferential theory of vision. On the other hand, Hering (1861) argued for a physiological interpretation of rivalry and much of the dispute with Helmholtz surrounded the visibility of yellow from dichoptic combinations of red and green. This reflected the disputes between them over trichromatic and opponent process theories of colour vision (see Turner, 1994). Panum (1858) and Hering (1861) considered that binocular rivalry was physiological whereas Wheatstone and Helmholtz maintained that it was psychological. In his *Handbuch* Helmholtz discussed contour rivalry in some detail and also emphasised that changing, complex mixtures of the two stimuli tend to be visible most of the time with only occasional periods in which the stimulus in one eye alone dominates.

Panum (1858) introduced crossed gratings as stimuli for binocular rivalry: “The rivalry of contours is at its strongest if the different lines in the two images are as equal as possible with regard to thickness and light intensity” (p. 38). The crossed diagonal figure was used by Helmholtz to support his theory that rivalry is a psychological rather than a physiological process because he could control which stimulus was visible: “These experiments show that man possesses the faculty of perceiving the images in each eye separately, without being disturbed by those in the other eye, provided it is possible for him… to concentrate his whole attention on the objects in this one field. This is an important fact, because it signifies, that *the content of each separate field comes to consciousness without being fused with that of the other field by means of organic mechanisms;* and that, therefore, *the fusion of the two fields in one common image, when it does occur, is a psychic act*” (Helmholtz, 1925, p. 499). Three Plates in the *Handbuch* illustrated stimuli for binocular rivalry, one of which is shown in Figure 3.

In 1839, the year following the announcement of the stereoscope, photography was introduced to the public. Soon after, paired photographs from slightly different positions were taken that could be viewed in a stereoscope to yield apparent depth. It was also possible to present a positive image to one eye and a negative of the same image to the other. However, the initial studies of binocular lustre by Dove (1851) used drawings of geometrical figures either as black on white or white on black. Helmholtz examined this aspect of binocular vision which he called ‘stereoscopic lustre’ and stimuli inducing it were accorded the status of a Plate in his *Handbuch*.

IV. PERCEPTION

Helmholtz’s lasting influence in visual science has related to neither his physical rigour nor his observational precision, rather to his epistemology as it was enunciated in part three of the *Handbuch* (see Hatfield, 1990; Turner, 1994). The final section of the volume, reviewing theories of vision, posed him the greatest problems and required the most protracted preparation. Although little he wrote on the issue was novel, he marshalled the arguments over a wider range of phenomena than others had done before. He summarized his position succinctly: “The sensations of the senses are tokens for our consciousness, it being left to our intelligence to learn how to comprehend their meaning" (Helmholtz, 1925, p. 533). The German text from Helmholtz (1896) for this carries Helmholtz's portrait in Figure 4. The text also emphasizes the influences that physicists like Wheatstone and Volkmann have had in developing his psychological theory of perception.

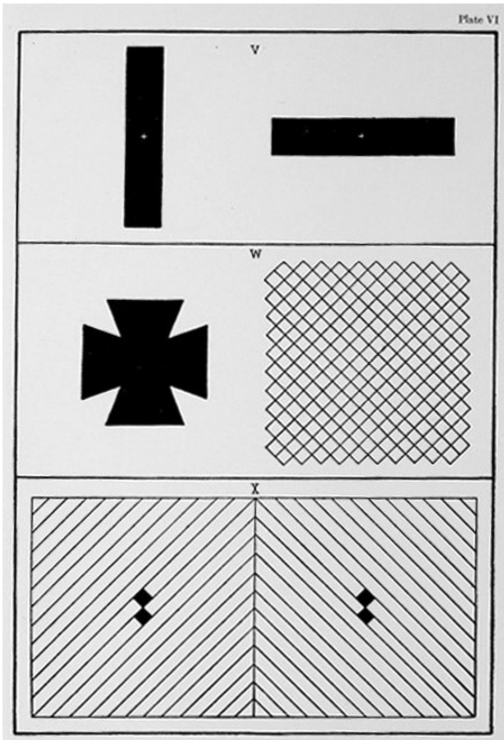


Figure 3. Left and right eye stimuli for binocular contour rivalry (Plate 6 from Helmholtz, 1925).

In developing the theory that sensory signs or tokens rather than images are the basis for perception, Helmholtz was guided by Müller's doctrine of specific nerve energies (see Cassedy, 2008; Finger & Wade, 2002). Müller (1826) argued that the particular sensations experienced are dependent on the nerves excited, no matter how those nerves are stimulated. For Helmholtz these sensory signs were processed rapidly, unconsciously and inferentially in what came to be called 'unconscious inferences'. A similar concept was advanced by Berkeley (1709) and it was employed by Wheatstone (1838) to interpret stereoscopic phenomena. Helmholtz's many debts to Wheatstone were acknowledged in his *Handbuch*. Not only did Wheatstone provide the instrument with which Helmholtz would challenge the nativism of Hering, he also supplied the theoretical framework, including the concept of unconscious mental processes that were used to support the stereoscopic observations. Wheatstone effectively reunited binocularity with space perception and he employed Berkeleyian empiricism to cement the union. Space perception had remained the province of philosophers for centuries, but it was clear to both Wheatstone and Helmholtz that the methods of physics could be applied in an effective way to study space and depth perception

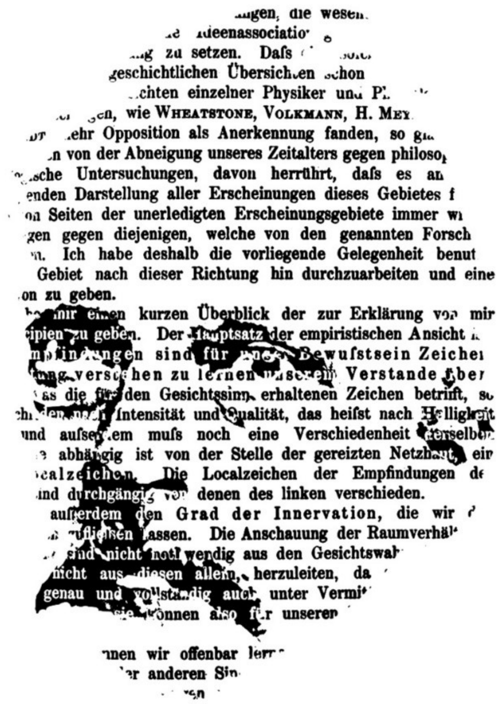


Figure 4. *Tokens of sensation (Hermann von Helmholtz)* by Nicholas Wade.

The image based theories had long been dominant following Kepler's (1604) analysis of ocular dioptrics: an inverted and reversed image of external objects was focused on the back of the eye. Kepler referred to it as a picture and set in train an approach to vision which involved processing this picture. This created problems like relating upright vision to an inverted image. As was the case with two retinal images, such problems were bypassed when considering sensory signs: "These two difficulties do not apply to the Empirical Theory, since it only supposes that the actual sensible 'sign,' whether it be simple or complex, is recognised as the sign of that which it signifies. An uninstructed person is as sure as possible of the notions he derives from his eyesight, without ever knowing that he has two retinae, that there is an inverted picture on each, or that there is such a thing as an optic nerve to be excited, or a brain to receive the impression. He is not troubled by his retinal images being inverted and double. He knows what impression such and such an object in such and such a position makes on him through his eyesight, and governs himself accordingly. But the possibility of learning the signification of the local signs which belong to our sensations of sight, so as to be able to recognise the actual relations which they denote, depends, first, on our having movable parts of our own body within sight; so that, when we once know by means of touch what relation in space and what movement is, we can further learn what changes in the impressions on the eye correspond to the voluntary movements of a hand which we can see. In the second place, when we move our eyes while looking at a field of vision filled with objects at rest, the retina, as it moves, changes its relation to the almost unchanged position of the retinal picture. We thus learn what impression the same object makes upon different parts of the retina" (Helmholtz, 1873, p. 278). After 1867 Helmholtz did continue to give popular lectures on vision and, although he did not acquire any further evidence for his empiricism, his opposition to nativism was undiminished.

V CONCLUSION

The age old problem of how we see the world as single with two eyes was seemingly solved by linking the Vieth-Müller circle to identical retinal points only for this to be challenged by the demonstration of stereoscopic depth perception. Opposition to Wheatstone's empiricist interpretation was almost unanimous within German science but Helmholtz provided support for it in his *Handbuch*. Helmholtz applied the methods of physics to studies of binocular vision but he did not adopt the procedures of Fechner's psychophysics. Rather, Helmholtz relied on those who did, like Volkmann.

References

Berkeley, G. (1709). *An essay toward a new theory of vision*. Dublin: Pepyat.

Brewster, D. (1844). On the law of visible position in single and binocular vision, and on the representation of solid figures by the union of dissimilar plane pictures on the retina. *Transactions of the Royal Society of Edinburgh*, 15, 349–368.

Brewster, D. (1849). Account of a new stereoscope. *Report of the British Association for the Advancement of Science. Transactions of the Sections*, pp. 6–7.

Brücke, E. (1841). Ueber die stereoscopischen Erscheinungen und Wheatstone's Angriff auf die Lehre von denidentischen Stellen der Netzhäute. *Archiv für Anatomie und Physiologie* 459-76

Cassedy, S. (2008). A history of the concept of stimulus and the role it played in the neurosciences. *Journal of the History of the Neurosciences,* 17, 405-432.

Dove, H. W. (1841). Die Combination der Eindrücke beider Ohren und beider Augen zu einem Eindruck. *Monatsberichte der Berliner preussische Akademie der Wissenschaften,* 41, 251-252.

Dove, H. W. (1851). Ueber die Ursache des Glanzes und der Irradiation, abgeleitet aus chromatischen Versuchen mit der Stereoskop. *Annalen der Physik und Chemie,* 83, 169-183.

Fechner, G. T. (1860). *Elemente der Psychophysik. Erster Theil*. Leipzig: Hirzel.

Fechner, G. T. (1861). Über einige Verhältnisse des binocularen Sehens. *Abhandlungen der mathematisch-physischen Classe der königlich sächsischen Gesellschaft der Wissenschaften*, 50, 344-564.

Finger, S., and Wade, N. J. (2002). The neuroscience of Helmholtz and the theories of Johannes Müller. Part 2. Sensation and perception. *Journal of the History of the Neurosciences,* 11,234-254.

Hatfield, G. (1990). *The natural and the normative: Theories of spatial perception from Kant to Helmholtz*. Cambridge, MA: MIT Press.

Helmholtz, H. (1857). Das Telestereoskop. *Annalen der Physik und Chemie,* 102,167-175

Helmholtz, H. (1867). Handbuch der physiologischen Optik. In *Allgemeine Encyklopädie der Physik* vol. 9. G. Karsten, ed. Leipzig: Voss

Helmholtz, H. (1873). *Popular lectures on scientific subjects*. First series. Trans. E. Atkinson. London: Longmans, Green.

Helmholtz, H. (1881). *Popular lectures on scientific subjects*. Second series. Trans. E. Atkinson. London: Longmans, Green.

Helmholtz, H. (1896). *Handbuch der physiologischen Optik.* 2nd ed. Leipzig: Voss.

Helmholtz, H. von (1924). *Helmholtz's Treatise on physiological optics*, vol. 1. Trans. J. P. C. Southall. New York: Optical Society of America.

Helmholtz, H. von (1925). *Helmholtz's Treatise on physiological optics*, vol. 3. Trans. J. P. C. Southall. New York: Optical Society of America.

Hering, E. (1861). *Beiträge zur Physiologie. I. Vom Ortsinne der Netzhaut*. Leipzig: Engelmann.

James, W. (1890). *Principles of Psychology*, vol. 2. London: Macmillan.

Kepler, J. (1604). *Ad Vitellionem paralipomena*. Frankfurt: Marinium and Aubrii.

Lenoir, T. (1993). The eye as mathematician. In *Hermann von Helmholtz and the foundations of nineteenth-century science,* ed. D. Cahan, 109-153. Berkeley: University of California Press.

Müller, J. (1826). *Zur vergleichenden Physiologie des Gesichtssinnes des Menschen und der Thiere, nebst einen Versuch über die Bewegung der Augen und über den menschlichen Blick*. Leipzig: Cnobloch.

Müller, J. (1838). *Handbuch der Physiologie des Menschen für Vorlesungen,* vol. 2. Coblenz: Hölscher

Ono, H. (1981). On Wells' 1792 law of visual direction, *Perception & Psychophysics*, 30, 403-406.

Panum, P. L. (1858). *Physiologische Untersuchungen über das Sehen mit Zwei Augen*. Kiel: Schwersche Buchhandlung.

Rollmann, W. (1853). Zwei neue stereoskopische Methoden. *Annalen der Physik und Chemie*, 166, 186–187.

Turner, R. S. (1994). *In the eye's mind. Vision and the Helmholtz-Hering controversy*. Princeton: Princeton University Press.

Vieth, G. U. A. (1818). Ueber die Richtung der Augen. *Annalen der Physik,* 28, 233-253

Volkmann, A. W. (1859a). Die stereoskopischen Erscheinungen in ihrer Beziehung zu der Lehre von den identischen Netzhautpunkten. *Archiv für Ophthalmologie*, 5, 1–100.

Volkmann, A. W. (1859b). Das Tachistoscop, ein Instrument, welches bei Untersuchung des momentanen Sehens den Gebrauch des elektischen Funkens ersetz. *Berichte über die Verhandlungen der königlichen sächsischen Akademie der Wissenschaften mathematische-naturwissenschaftliche Klasse,* 2, 90-98.

Wade, N. J. (1983). *Brewster and Wheatstone on vision*. Academic Press: London.

Wade, N. J. (1998). *A natural history of vision.* MIT Press: Cambridge, MA.

Wade, N. J. (2019). Ocular equivocation. The rivalry between Wheatstone and Brewster. *Vision*, 3(2), 26.
Wade, N. J. (2021a). On the origins of terms in binocular vision. *i-Perception* 12(1):1–19.
Wade, N. J. (2021b). Helmholtz at 200. *i-Perception* 12(4):1–19.
Wade, N. J. and Ngo, T. (2013). Early views on binocular rivalry. In *Constitution of visual consciousness: Lessons from binocular rivalry* ed. S. Miller, 77-108. Amsterdam: John Benjamins.
Wade, N. J., and Ono, H. (2012). Early studies of binocular and stereoscopic vision. *Japanese Psychological Research,* 54, 54-70.
Wells, W. C. (1792). *An essay upon single vision with two eyes: Together with experiments and observations on several other subjects in optics*. London: Cadell.
Wheatstone, C. (1838). Contributions to the physiology of vision - Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision, *Philosophical Transactions of the Royal Society,* 128, 371-394.
Wheatstone, C. (1839). Ueber das Sehen mit zwei Augen und das Stereoskop. *Annalen der Physik und Chemie*, 47, 625-627.
Wheatstone, C. (1842). Beiträge zur Physiologie des Gesichtssinnes. Erster Teil. *Annalen der Physik und Chemie,* Supplementary volume 1, 1-48.

The Dynamics of Perceptual Filling-in and Motion Induced Blindness during Binocular Rivalry: Exploring the Common Mechanism Hypothesis

Rebecca A. White,[1] John E. Sparrow,[2] Kaylyn Dickinson,[1] Danielle Tremblay,[1]
Joseph A. LaBarre,[1] Andrew J. Kitt,[1] and Wm Wren Stine[1]

[1]Department of Psychology, University of New Hampshire, Durham, NH, USA
[2]Social Science Division, University of New Hampshire-Manchester, Manchester, NH, USA
raw1033@wildcats.unh.edu

Abstract — Motion induced blindness (MIB) and perceptual filling-in (PFI) are phenomena wherein the visibility of a target stimulus fluctuates when presented upon a globally changing pattern. Similarities in the temporal dynamics of these phenomena and that of binocular rivalry (BR) have led researchers to hypothesize that they are all forms of perceptual rivalry that share a single common oscillatory mechanism, though this hypothesis has been challenged by counterevidence from Jaworska and Lages' (2014). In the present study, the dynamics of MIB and PFI were examined during BR and during identical conditions that did not induce BR. Results indicated no effect of BR on MIB. However, BR significantly increased all measures of PFI. Furthermore, MIB and PFI dynamics, when compared to each other, expressed marked differences, regardless of BR. These results are counterintuitive given the current literature, which goes so far as to suggest that MIB and PFI are in fact the same phenomena under differing perceptual conditions (e.g., New & Scholl, 2008).

Keywords – Motion Induced Blindness, Perceptual Filling-in, Binocular Rivalry, Common Mechanism Hypothesis

I. INTRODUCTION

Motion induced blindness (MIB) is a phenomenon wherein a salient, peripheral target stimulus vanishes from conscious visual awareness, often for several seconds at a time, when presented with global changes (i.e., a "motion mask"; Bonneh et al., 2001). Current understanding of the mechanisms and processes driving MIB is limited. Evidence has been provided for several different theories explaining MIB (e.g., motion streak suppression; Wallis & Arnold, 2009), but these theories do not provide a full account for the multiple neural areas which MIB activates within the visual hierarchy, including both high- and low-level regions (Donner et al., 2008, 2013).

Perceptual filling-in (PFI) is another phenomenon which renders target stimuli perceptually invisible (Anstis, 1989; Ramachandran et al., 1993). Like MIB, PFI occurs when a target stimulus is superimposed with global motion, but rather than a target which is suppressed, a gap within the stimuli is "filled in" by its surrounding pattern (Ramachandran & Gregory, 1991). New and Scholl (2008, 2018) hypothesize that MIB and PFI are functional responses by the visual system to an artificial scotoma, such that the targets are misjudged as a disruption within the visual system itself. This hypothesis suggests that MIB and PFI are in fact the same phenomenon – a mechanism for filtering irrelevant and obstructing proximal percepts such as scotomas, where the optic nerve connects to the retina ("the blind spot"), and pieces of detached retina ("floaters"). Hsu and colleagues (2004, 2006) provide further support, demonstrating that the few differences between MIB and PFI in the literature at that time (e.g., less filling in compared to MIB) are no longer observed when eccentricity, contrast, and perceptual grouping are accounted for in the stimulus display.

Binocular rivalry (BR) is characterized by stochastic oscillations in conscious awareness between two sufficiently different (i.e., unfusbale) images presented to each eye (von Helmholtz, 1925). Another well documented interpretation of MIB suggests it constitutes a global versus local form of perceptual rivalry, sharing a common mechanism with BR (Carter & Pettigrew, 2003; Doner, Sagi, & Bonnet, 2008). Carter and Pettigrew (2003) revealed that the temporal dynamics between the disappearance and reappearance in MIB and the switch rate between perceptually masked images during BR are remarkably similar within individuals, and they suggest that the two phenomena could be modulated by a common oscillatory pattern expressed by the pontine brainstem. BR switch rates and MIB disappearance/re-appearance rates are also identically altered via Transcranial Magnetic Stimulation (TMS; Funk & Pettigrew, 2003), administration of certain psychotropic drugs (Carter & Pettigrew, 2003; Carter et al., 2005, 2007), and in clinical populations, such as those with psychosis, OCD, or autism (Robertson et al., 2013; Tschacher et al., 2006; Vierck et al., 2013; Ye et al., 2013). While these findings agree with a common mechanism hypothesis, they do not necessarily confirm or deny the idea. Jaworska and Lages (2014) provided counterevidence by dichoptically presenting targets rivalrous in color and inducing MIB with a motion mask, finding that certain MIB manipulations had no effect on the dynamics of BR, and vice versa, suggesting that MIB and BR involve different processing when competing for visual awareness (Blake & Logothetis, 2002; Sterzer et al., 2009). However, the mechanisms driving color rivalry have been shown to be separate from those which drive rivalries between forms (Holmes et al., 2005) as well as motion and form (Carney et al., 1987).

In the present study, MIB and PFI dynamics were examined during BR and matched control stimuli. If these phenomena are driven by a single common mechanism, there should be consistent similarities in the temporal dynamics (e.g., the frequency and duration of perceptual reversals) within an individual, and thus these dynamics should change during BR as compared to a control condition (a non-rivalrous MIB/PFI display).

II. METHODS

Participants. Seven volunteers with normal or corrected to normal vision were recruited. Participants signed informed consent forms and received debriefing consistent with University of New Hampshire Institutional Review Board policy. There was no compensation.

Apparatus & Stimuli. Stimuli were presented through a Maxwellian View optical system (Beer et al., 2005; Westheimer, 1966) with 3.0 mm artificial pupils and produced by a 27" 5120 x 2880 pixel iMac Pro monitor with a 60 Hz refresh rate, CIE coordinates for white at 0.3127 and 0.3290, and color display support of 2^{10} levels per RGB. Luminance was γ-corrected, ranging from 10 to 350 cd/m^2, or 54.5 to 1905 Trolands (Td) of retinal illuminance, given the 3.0 mm apertures and 77% transmittance via the optical system. Rapid adjustment of interocular distance was able to be performed using a modified tank periscope with right angle prisms in each optical path of the viewing system. Including the path of the prisms, the apertures were at a total distance of 1664 mm from the screen. A vertical partition was fixed along the participant's line of sight to separate the sets of stimuli for each eye. Stimuli were placed in the plane of the artificial pupils, passing through rhomboid prisms and two lenses. The first set of lenses are at one focal length (defined by light rays being columnated) and the second set are two focal lengths from the first set, including the path through the rhomboid prisms between them.

Stimuli were created in *MatLab* (v. 2019b, MathWorks) using the *PsychToolbox* extensions (Brainard, 1997; Pelli, 1997) and consist of nonius lines, a grid with a hole in the center, opposing lines, and 64 randomly placed moving dots which comprise the motion mask. The grid, opposing lines, and nonius lines are white and at maximum positive contrast with respect to a grey background (170 cd/m^2; 979.75 Td), while the motion mask is black and at maximum negative contrast. Both the maximum positive and negative is +/- 94.4% Weber contrast. The moving dots of the mask (in addition to all other stimuli) are contained in squares with side lengths subtending 2.2º and centered over the nonius lines, which are centered for each eye. Nonius lines each span 5 arcmin. After a dot travels off the edge of this square, it re-appears on the opposite end and continues in the same direction (i.e., it wraps around). Individual mask dots are 1 arcmin. in diameter and move at a speed of 4º per second incoherently in random directions, given that incoherent motion is more effective at inducing MIB than coherent motion (LaBarre & Stine, 2019; Wells et al., 2011). The grid is a square with side lengths of .80º, spacing of 9.6 arcmin., 2.0 arcmin. wide bars, and a 10 arcmin. section removed from the second bar. The opposing lines are identical in size and eccentricity (.90º from the nonius lines) to the missing section of the grid, ensuring that the area undergoing MIB/PFI is consistent across all trials. All relevant combinations of these stimuli were tested, giving 6 total sets of stimuli.

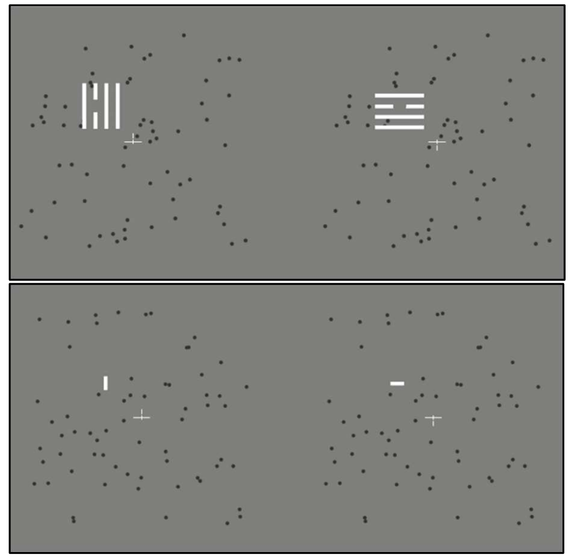


Figure 1. Top: rivalrous PFI stimuli. Bottom: rivalrous MIB stimuli.

Procedure. Trials took place in a darkened room after a 120 second adaptation period. Participants used a bite bar for head stabilization while viewing the stimuli through the Maxwellian view system. Each set of the six stimuli were presented in a randomized order eight times, with trials that lasted 30 seconds. Trials were preceded by a 30 second adaptation period to avoid confounding aftereffects. Thus, the experiment took approximately one hour for each participant to complete, with the option to take a break half-way through. Regardless of whether the option to take a break was chosen, the second half of the experiment began with another 120 second adaptation period. To control for effects of ocular dominance, the rivalrous stimuli were programmed such that there was an equal probability for the horizontal and vertical lines to be presented in either eye on any given trial, with the constraint that there was an equal number of presentations of the horizontal and vertical lines in each eye. Participants were instructed to indicate MIB and PFI effects by pressing spacebar on a keyboard throughout the duration of target disappearance or filling in, respectively, for each trial. Key presses were sampled at a rate of 60 Hz.

III. RESULTS

The total normalized MIB/PFI effect as indicated by key press data was analyzed using a factorial subject by stimulus (6 levels) and trial (8 levels) repeated measures analysis of variance (randomized block factorial design, RBF-68; Kirk, 2013, Ch. 10). This model was constructed for three dependent variables: cumulative duration of the effect, the duration to the first onset of disappearance/filling in, and perceptual reversals (i.e., how many times the target stimulus came into or out of awareness). Thus, a critical alpha level of .016 was used to control the familywise Type I error rate at .05. An alpha level of .15 was used for exploratory data analyses. According to the locally best invariant test and Tukey's test of nonadditivity, the sphericity and additivity assumptions, respectively, were violated across all three measures. Thus, all critical F and p values were adjusted using the Chi Muller statistic and the block by treatment interactions were explored. An arcsine transform was used to correct for violations of normality in the total duration and duration to onset. Given the exploratory nature of this experiment, the *a priori* decision was made to test the omnibus hypothesis that our stimuli differentially influenced each of our response measures. Significant omnibus tests were followed by *a posteriori* pairwise comparisons using the Simulate adjustment procedure (*SAS* v9.4), which controls the familywise Type I error rate at .05 using the control variate adjustment method of Hsu and Nelson (1998).

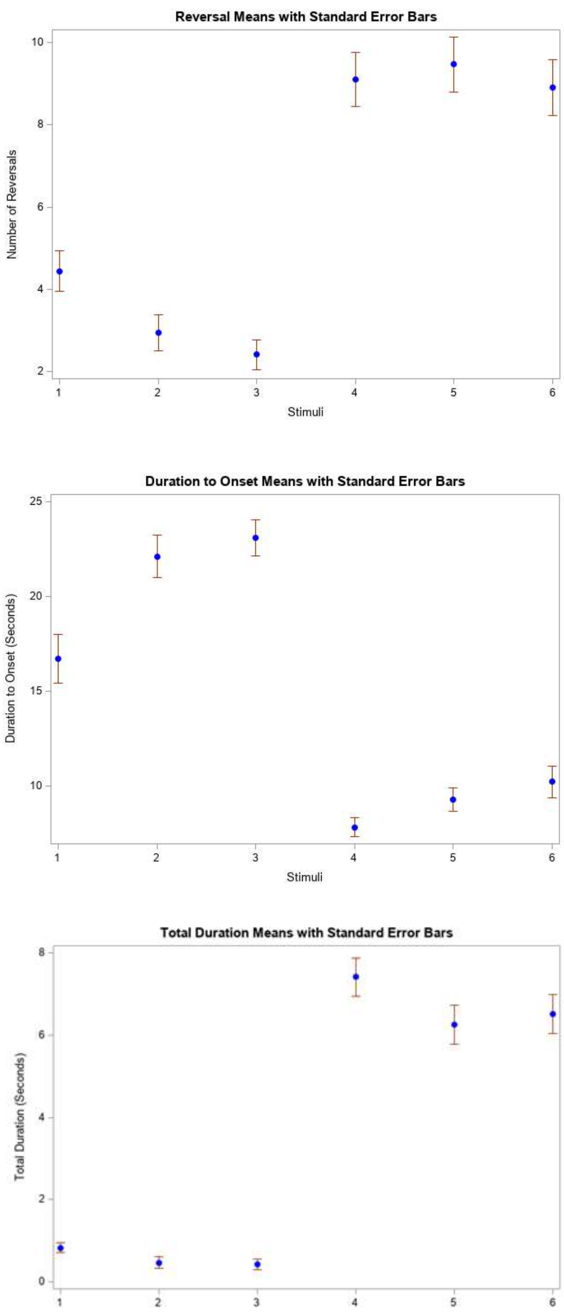


Figure 2. Means and standard errors. Stimuli 1-3 represent PFI conditions, with 1 inducing BR. Stimuli 4-6 represent MIB conditions, with 4 inducing BR.

There was a significant effect of stimuli on total duration, *CM Adj.* $F(0.4, 2.4) = 277.55$, $p = .002$; partial $\omega^2 = 0.83$; duration to onset, *CM Adj.* $F(1.3, 7.8) = 78.88$, $p = 1.44\text{x}10^{-5}$; partial $\omega^2 = 0.54$; and reversals, *CM Adj* $F(2.95, 14.75) = 107.68$, $p = 3.15\text{x}10^{-10}$; partial $\omega^2 = 0.61$. Pairwise comparisons using the Simulate procedure revealed that all three dependent variables were significantly greater for MIB than PFI stimuli (all *p's* < .0001). The BR PFI stimuli induced significantly increased effects on all three measures when compared to the nonrivalrous horizontal and vertical grids. There were no significant differences between the rivalrous and non-rivalrous conditions in the MIB stimuli, and the non-rivalrous horizontal and vertical controls did not differ in either the MIB or PFI conditions.

Use of the subject blocking variable was effective for total duration, *CM Adj.* $F(.48, 16.8) = 42.12$, $p = 8.79\text{x}10^{-5}$; $\rho = .99$, duration to onset, *CM Adj.* $F(1.56, 54.6) = 17.51$, $p = 7.89\text{x}10^{-6}$; $\rho = .99$, and reversals $F(3.54, 123.9) = 86.82$, $p < 1.00\text{x}10^{-16}$, $\rho = .99$. There was no effect of trial.

IV. DISCUSSION

The present study aimed to investigate MIB and PFI during BR to explore whether or not they share a common mechanism, measuring the dynamics of these phenomena on three measures. The reversal rate, onset to the first disappearance, and total duration of disappearance in MIB did not change when presented with rivalrous forms as opposed to non-rivalrous forms. For PFI, all three of these measures were significantly altered. Furthermore, there was less filling in than disappearance overall, regardless of BR. Th ese findings are counter-intuitive to our current understanding of MIB and PFI, given that the target region in the displays were matched in terms of size, contrast, and eccentricity (Hsu et al., 2004, 2006). It should be noted that, although these PFI stimuli were derived from New and Scholl's (2008) experiment, where a hole was removed from a gridded texture, the approach of the present study is novel in that PFI is arguably being induced via the Gestalt principle of good continuation. Thus, the filling in that occurred may not have the same dynamics and/or is not modulated by the same mechanisms as PFI induced by conventional displays. Further research is needed.

The original PFI display consisted of a finely gridded texture similar to that of New and Scholl (2008). However, pilot testing revealed that these stimuli fuse rather than inducing BR. Spacing of the individual grid lines as well as their thickness play an important role in determining a threshold between fusion and BR. Thus, another confound of the PFI stimuli is that the visual system may have responded with MIB processes in addition to PFI processes, such that there was an attempt to mask individual grid lines. Again, further research is needed.

Outside of the PFI conditions, results from the present study further support Jaworska and Lages' (2014) conclusion that the common mechanism hypothesis may be incomplete. There may be some degree of interaction between the systems modulating these phenomena, and they likely share many computational principles, but not a single oscillatory mechanism.

References

Alais, D., Cass, J., O'Shea, R. P., & Blake, R. (2010). Visual sensitivity underlying changes in visual consciousness. *Current Biology*: CB, 20, 1362‑1367. https://doi:10.1016/j.cub.2010.06.015

Beer, R. D., Macleod, D. I., & Miller, T. P. (2005). The Extended Maxwellian View (BIGMAX): A high-intensity, high-saturation color display for clinical diagnosis and vision research. *Behavior Research Methods,* *37*(3), 513-521. https://doi:10.3758/bf03192722

Blake, R., & Logothetis, N. K. (2002). Visual Competition. *Nature Reviews Neuroscience*, 3, 13–21. https://doi:10.1038/nrn701

Bonneh, Y. S., Cooperman, A., & Sagi, D. (2001). Motion-induced blindness in normal observers. *Nature*, *411*(6839), 798–801. https://doi.org/10.1038/35081073

Brainard, D. H. (1997). The psychophysics toolbox. Spatial Vision, 10, 433–436.

Carney, T., Shadlen, M., & Switkes, E. (1987). Parallel processing of motion and colour information. *Nature, 328*(6131), 647-649. https://doi:10.1038/328647a0

Carter, O. L., & Pettigrew, J. D. (2003). A common oscillator for perceptual rivalries? *Perception, 32*(3), 295-305. https://doi:10.1068/p3472

Carter, O. L., Pettigrew, J. D., Hasler, F., Wallis, G. M., Liu, G. B., Hell, D., & Vollenweider, F. X. (2005). Modulating the rate and rhythmicity of perceptual rivalry alternations with the mixed 5-HT2A and 5-HT1A agonist Psilocybin. *Neuropsychopharmacology, 30*, 1154-1162. https://doi:10.1038/sj.npp.1300621

Carter, O. L., Hasler, F., Pettigrew, J. D., Wallis, G. M., Liu, G. B., & Vollenweider, F. X. (2007). Psilocybin links binocular rivalry switch rate to attention and subjective arousal levels in humans. *Psychopharmacology,* (195), 415-424. https://doi:10.1007/s00213-007-0930-9

Devyatko, D., Appelbaum, L. G., & Mitroff, S. R. (2016). A common mechanism for perceptual reversals in motion-induced blindness, the Troxler effect, and perceptual filling-in. *Perception, 46*(1), 50-77. https://doi:10.1177/0301006616672577

Edwards, D., & Berry, J. J. (1987). The efficiency of simulation-based multiple comparisons. *Biometrics, 43*(4), 913. https://doi:10.2307/2531545

Funk, A. P., & Pettigrew, J. D. (2003). Does interhemispheric competition mediate motion-induced blindness? A transcranial magnetic stimulation study. *Perception, 32*(11), 1328-1338. https://doi:10.1068/p5088

Holmes, D. J., Hancock, S., & Andrews, T. J. (2006). Independent binocular integration for form and colour. *Vision Research, 46*(5), 665-677. https://doi:10.1016/j.visres.2005.05.023

Hsu, J. C., & Nelson, B. (1998). Multiple comparisons in the General Linear Model. *Journal of Computational and Graphical Statistics, 7*(1), 23-41. https://doi:10.1080/10618600.1998.10474759

Hsu, L.C., Yeh, S.-L., & Kramer, P. (2004). Linking motion-induced blindness to perceptual filling-in. *Vision Research*, 44, 2857 ‑ 2866. https://doi:10.1016/j.visres.2003.10.029

Hsu, L.C., Yeh, S.-L., & Kramer, P. (2006). A common mechanism for perceptual filling-in and motion-induced blindness. *Vision Research*, 46, 1973‑1981. https://doi:10.1016/j.visres.2005.11.004

Jaworska, K., & Lages, M. (2014). Fluctuations of visual awareness: Combining motion-induced blindness with binocular rivalry. *Journal of Vision, 14*(11), 11. https://doi:10.1167/14.11.11

Kirk, R. W. (2013). Experimental design: Procedures for the behavioral sciences (4th ed.). Washington, DC: Sage.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.

New, J. J., & Scholl, B. J. (2008). "Perceptual Scotomas." *Psychological Science*, *19*(7), 653–659. https://doi.org/10.1111/j.1467-9280.2008.02139.x

New, J. J., & Scholl, B. J. (2018). Motion-induced blindness for dynamic targets: Further explorations of the perceptual scotoma hypothesis. *Journal of Vision*, *18*(9), 24. https://doi.org/10.1167/18.9.24

Ramachandran, V. S., & Gregory, R. L. (1991). Perceptual filling in of artificially induced scotomas in human vision. *Nature, 350*(6320), 699-702. https://doi:10.1038/350699a0

Robertson, C. E., Kravitz, D. J., Freyberg, J., Baron-Cohen, S., & Baker, C. I. (2013). Slower rate of binocular rivalry in autism. *The Journal of Neuroscience, 33*(43), 16983-16991. https://doi:10.1523/jneurosci.0448-13.2013

Wallis, T. S., & Arnold, D. H. (2009). Motion-Induced Blindness and Motion Streak Suppression. *Current Biology, 19*(4), 325-329. https://doi:10.1016/j.cub.2008.12.053

Wells, E. T., Leber, A. B., & Sparrow, J. E. (2011). The role of mask coherence in motion-induced blindness. *Perception, 40*, 1503–1518. https://doi:10.1068/p6976

Westheimer, G. (1966). The Maxwellian view. *Vision Research, 6*(11-12), 669-682. https://doi:10.1016/0042-6989(66)90078-2

Ye, X., Zhu, R., Zhou, X., He, S., & Wang, K. (2019). Slower and less variable binocular rivalry rates in patients with bipolar disorder, OCD, major depression, and schizophrenia. *Frontiers in Neuroscience, 13*. https://doi:10.3389/fnins.2019.00514

Eye Movement Abnormalities among Patients with Schizophrenia

Alexandra Wolf, [1] Kazuo Ueda,[2] Yoji Hirano[3]

[1]Department of Human Science/Research Center for Applied Perceptual Science, Kyushu University, Fukuoka, Japan
[2]Unit of Perceptual Psychology, Department of Human Science/Research Center for Applied Perceptual Science/Division of Auditory and Visual Perception Research, Research and Development Center for Five-Sense Devices, Kyushu University, Fukuoka, Japan
[3]Department of Neuropsychiatry, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan
olaokami@gmail.com

Abstract— Neurophysiological investigations have shown that information processing among patients with first-episode and chronic schizophrenia (SZ) is abnormal and probably related to the core pathology of SZ. Therefore, SZ is ultimately about how the patient's brain processes information. In recent years, with the help of modern eye-tracking technology, scientists can recognize cognitive deficits among clinical populations more precisely. For example, while visually exploring a picture, SZ patients show restricted scanning patterns characterized by smaller eye movements (saccades) and fewer eye fixations, featuring returns to already visually examined regions. For instance, if it is true that 'prevention is better than the cure' it is necessary to undertake a detailed investigation of eye movement anomalies and draw scientists' attention to explore visual information processing patterns among SZ patients. This work undertakes a brief review of research on exploratory eye movement deficits among patients with SZ. To understand the relationship between eye movement impairments of patients with SZ as well as what tests are capable of identifying these abnormalities, the authors gather promising studies that have mounted in recent years. Furthermore, gaze metrics that identify the process of SZ are discussed in this work. Given that cognitive impairments appear before the official diagnosis is made, recognizing individuals in the early phases of SZ is crucial. Notably, eye-tracking technology has the potential to contribute to the process of early intervention.

Keywords— schizophrenia, eye movements, eye-tracking, diagnosis, abnormalities

IV. INTRODUCTION

Already at the end of the XIX century, experimental psychologists hypothesized that eye movements disclose processes that take place inside the mind of an observer. Since the retina develops from the same tissue as the brain, the concept of *"the eyes being the window to the mind"* intrigued researchers from interdisciplinary fields of science (e.g., philosophy, psychology, linguistics). By merely observing the eyes, D'Lamare indicated that human gaze behavior is not linear and that one's eyes are in perpetual motion, i.e., *"they fix a point, make a movement, fix another point, and so forth"* [1].

Next, ground-breaking investigations with the use of a non-invasive apparatus drew attention to the effect of instructions on the pattern of eye movements (kindly refer to Fig.1) [2]. Later on, the knowledge on the relationship between eye movements and the sequence of thought (information processing strategy) has been extended by Yarbus, who presented qualitative data, showing that the viewers' eyes are directed towards areas of the image that are *"useful or essential"* to perception [3].

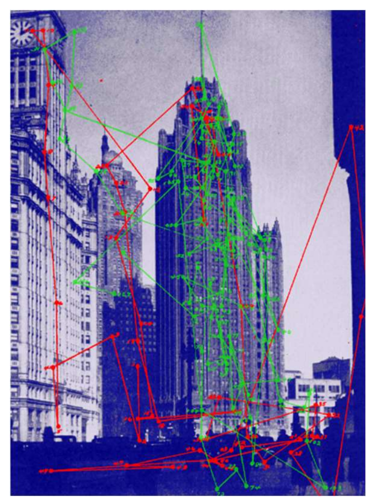


Figure 1. Eye trajectories from one subject during two free-viewing tasks; one without *(red)* and one with an instruction *(green)* to find a person looking out of a window. Obtained from Buswell, 1935.

V. GAZE FORMULATES PSYCHOLOGICAL LAWS

In modern times, thanks to sophisticated mathematical image analysis, eye-trackers stand out for providing replicable, unbiased, and reliable behavioral data. Gaze metrics such as duration of viewing time, scan-path length, or the number of blinks, provide quantitative insights into human patterns of thinking, believes, and behavior [4]. Despite individual differences in visual dynamic (e.g., the sequence of eye

movement), the human gaze provides sufficient parameters to formulate psychological laws.
Wolf and Ueda (2021) suggested that in clinical research, eye tracking can be a useful tool to objectively explore deviations in scanning patterns that account for cognitive impairments (deficits in intellectual functioning, i.e., attention, memory and executive function) associated with psychological disorders and personality entities [5].

Although, it has been suggested that exploratory gaze behavior in clinical populations differs considerably from that among healthy controls, only few investigations undertook to study and characterize the exploratory eye movements among clinical populations. For example, in 1972, by showing a horizontal 'S' shaped figure to chronic SZ patients, Moriya *et al.* found that SZ patients made fewer movements in comparison with healthy controls. In addition, the range of eye movements among SZ patients was narrower. The researchers concluded that exploratory eye movement dysfunction reflects the trait of SZ.

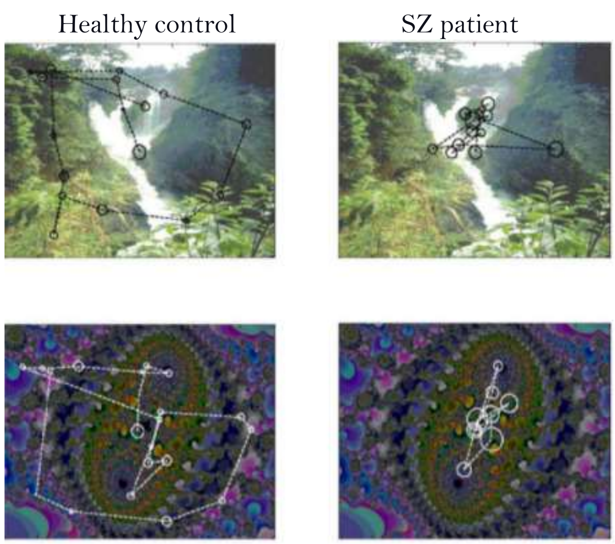


Figure 2. Illustrative examples of different eye movements between healthy control and a schizophrenia (SZ) patient viewing a picture (top: landscape; bottom: fractals). Reproduced from Bestelmayer *et al*. (2006).

A modern study of Bestelmayer *et al.,* where the participants were asked to freely view various types of images, replicated the previous finding of a global scanning impairment (Fig.2). Notably, the restricted visual scanning effect was present in all types of used stimuli (landscape, fractal, and faces) [6]. In summary, patients with schizophrenia exhibit focal processing strategy, whereas healthy controls efficiently minimize the global uncertainty by visually exploring the environment more widely (i.e., ambient processing) [7].

Given examples show that exploratory eye movement dysfunction appears to be specific to schizophrenia spectrum disorder. However, this mental illness is considered to be a constellation of several biologically different conditions [8]. Hence, the clinical community cannot expect to cover all of the conditions by a single viewing test, following an important note made by Matsushima *et al*., '(…) *exploratory eye movements alone cannot be used to pick up all schizophrenics'*. (p. 294) [9]. A combination of gaze metrics obtained by multiple tasks, however, may increase the classification accuracy to distinguish patients with schizophrenia from healthy controls (see Fig.3).

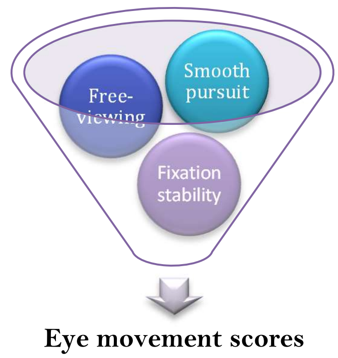


Figure 3. Combination of eye movement tests to develop eye movement scores. Gaze parameters from these tests, e.g., scan path length, fixation number, viewing time (from the free-viewing test); horizontal position gain (from the smooth pursuit test); fixation duration (from the fixation stability test) are being used to differentiate SZ patients from the healthy population.

Finally, several imaging studies have mentioned an overlap of the impaired brain circuit among schizophrenia patients with that of individuals with autism spectrum disorder. Both groups of patients share common cognitive symptoms; therefore, to state a precise diagnosis is arduous. This notion marked a new trend in psychiatry, i.e., to examine eye movements among adults with SZ and autism spectrum disorder, and compare the results with the performance of a typically developed (healthy) group [10].

Recently, Shiino *et al.* aimed to understand the similarities and differences in eye movement abnormalities across individuals diagnosed with schizophrenia and those with autism. The research group performed 3 viewing tests (i.e., free-viewing, smooth pursuit, and fixation stability). Out of 75 collected eye movements, they successfully pointed out five characteristics where patients with schizophrenia showed significant differences from the individuals with autism [11].

VI. FUTURE APPLICATION

Mentioned promising findings are just the beginning. As scientists continue to explore eye movement abnormalities among clinical populations, more accurate predictive tests for schizophrenia can emerge. Here, it is essential to mention that experimental paradigms influence what scientists accept as truth. Dowiasch *et al.* reported that some of the results obtained from a real-life study were contradictory to previous findings from indoor experimental settings [12]. Therefore, ecologically valid experiments should be investigated among the clinical populations as well.

Finally, cognitively informative paradigms are key to tackle psychiatric disorders [4], [5]. Yet, information processing strategies are scarcely being investigated in the clinical domain. Notably, this new trend in neurophysiological investigations may clarify the patterns of thinking, believes, and behavior among patients with psychiatric disorders.

References

[1] Tscherning, M. H. E. (1898) Optique physiologique, Paris, France: Masson and others, Booksellers of the Academy of Medicine.

[2] Buswell, G. T. (1935). How people look at pictures: a study of the psychology and perception in art. Univ. Chicago Press.

[3] Yarbus A. L. (1967). *Eye movements and vision.* New York: Plenum.

[4] Wolf, A., Ueda, K., and Hirano, Y. (2021). Recent updates of eye movement abnormalities in patients with schizophrenia: a scoping review. *Psychiatry Clin. Neurosci.* 75, 82–100. doi: 10.1111/pcn.13188

[5] Wolf, A. and Ueda, K. (2021) Contribution of Eye-Tracking to Study Cognitive Impairments Among Clinical Populations. *Front. Psychol.*12:590986. doi: 10.3389/fpsyg.2021.590986

[6] Bestelmeyer PEG, Tatler BW, Phillips LH, Fraser G, Benson PJ, Clair DS. (2006). Global visual scanning abnormalities in schizophrenia and bipolar disorder. Schizophr. Res. 87: 212–222.

[7] Suzuki M, Takahashi S, Matsushima E et al. (2009). Exploratory eye movement dysfunction as a discriminator for schizophrenia: A large sample study using a newly developed digital computerized system. Eur. Arch. Psychiatry Clin. Neurosci.259: 186–194.

[8] American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. Arlington: American Psychiatric Association; 2013

[9] Matsushima E, Kojima T, Ohta K et al. (1998). Exploratory eye movement dysfunction in patients with schizophrenia: Possibility as a discriminator for schizophrenia. J. Psychiatr. Res. 32: 289–295.

[10] Fukushima J. (2012). Neurophysiological studies in autism spectrum disorders—comparison with those in schizophrenia. Seishin Shinkeigaku Zasshi 114: 335–348 (in Japanese).

[11] Shiino T, Miura K, Fujimoto M et al. (2020). Comparison of eye movements in schizophrenia and autism spectrum disorder. Neuropsychopharmacol. Rep. 40: 92–95.

[12] Dowiasch S, Backasch B, Einhäuser W, Leube D, Kircher T, Bremmer F. (2016). Eye movements of patients with schizophrenia in a natural environment. *Eur. Arch. Psychiatry Clin. Neurosci.* 266: 43–54.

Fechner's Methods for Measuring Differential Sensitivity: Do the Results Correlate for Visual Tests of Line Length?

Nan Yang[1], Roger Adams[2], Jia Han[3,4,5] & Gordon Waddington[2]

[1]School of International Education, Shanghai University of Sport, Shanghai 200438, China
[2]Faculty of Health, University of Canberra, ACT 2601, Australia
[3]School of Kinesiology, Shanghai University of Sport, Shanghai 200438, China
[4]Research Institute for Sport and Exercise, University of Canberra, ACT 2601, Australia
[5] Faculty of Health, Arts and Design, Swinburne University of Technology, Hawthorn
Nan.Yang@canberra.edu.au; Roger.Adams@canberra.edu.au; Jia.Han@canberra.edu.au; Gordon.Waddington@canberra.edu.au

Abstract—Fechner's three psychophysical methods - the method of constant stimuli, the method of adjustment or average error (AE), and the method of limits - have been widely used in testing sensitivity, but the relationship between the results from the same individuals on these three methods is seldom investigated. The present study aimed to determine the relationship between sensitivity scores obtained from three test methods - Constant Stimuli, Absolute Judgement, and Adjustment - with visual stimuli. The same participants completed the three visual differential sensitivity tasks, tested with the same line lengths using different methods, with the stimulus lines shown on cards. The Adjustment Method AE scores were not significantly correlated with the Method of Constant Stimuli JNDs, the Absolute Judgement Method AUCs, or the Absolute Judgement AE scores. The only significant correlation between scores obtained from the different methods was found between the participants' JNDs for the Constant Stimuli Test and their AUCs for the Absolute Judgement Test, which were negatively correlated. The results are consistent with the view that, for both visual and proprioceptive sensory modalities, scores obtained with Fechner's experimental methods for measuring differential sensitivity applied to the same individuals do not tend to be correlated, and therefore do not preserve ranking of the individuals across the methods. We argue that this is because the methods depend on different cognitive processes underlying the performance of discrimination and reproduction tasks.

Keywords— Fechner, Method of Constant Limits, Method of Adjustment/Reproduction, Method of Absolute Judgement

I. INTRODUCTION

Fechner (1860) established three methods of measurement of differential sensitivity to stimulation - the method of right and wrong cases (later constant stimuli), the method of adjustment or average error, and the method of limits. He considered these three methods to complement each other in reaching the same goal (Elemente der Psychophysik, p.62). The relationship between scores from these three methods has not often been a focus for later studies. Silvestri and Fisischelli (1985), in their review of different interpretations of the three classical psychophysical methods, proposed a 'three faces' metaphor for the three methods of measuring sensitivity, seeing the methods as tasks being carried out by three different individuals – a sentry, a bombardier, and a diagnostician. The sentry needs sensitivity for detecting a stimulus that is different from the background (thus, Limits), the bombardier needs to adjust the signal to match a criterion value so as to have minimal error (Adjustment), and the diagnostician needs to identify the presented stimulus with the smallest frequency of error (Constant Stimuli).

For assessing proprioception, Fechner's three methods have subsequently been developed into the threshold to detection of passive motion (TTDPM) (Konczak et al. 2007), joint position reproduction (JPR), also known as joint position matching (Westlake et al. 2007; Goble 2010), and active movement extent discrimination assessment (AMEDA) testing (Waddington et al. 1999; Han 2013). If these methods are measuring the same attribute in the same individuals, then proprioceptive sensitivity scores should be correlated when obtained from TTDPM, JPR and discrimination tests. However, several studies have compared two of the methods for assessing joint proprioception, and concluded that each method examines different physiological aspects of proprioceptive function. That is, thresholds obtained from TTDPM tests correlated poorly with both position errors from JPR tests and with just noticeable difference (JND) scores from movement discrimination tests (de Jong et al. 2005; Elangovan et al. 2014; Nagai et al. 2016). Nagai et al. (2016) compared knee proprioception using TTDPM and active JPR methods, and also reported very weak correlations. In other studies, de Jong et al. (2005) and Elangovan et al. (2014) compared the movement detection threshold from TTDPM testing with a movement discrimination score, and neither of the correlation coefficients obtained in these studies were significant. In Yang et al (2018), the lack of correlation between scores from the three psychophysical methods was attributed to the methods requiring cognitive work at different levels, or 'depth of processing' (Craik & Lockhart, 1972). In a study by Yang et al. (2020), test results obtained from active joint position reproduction and active movement extent discrimination assessments by using the same apparatus, same testing requirement (active movement), and same scoring system (both using Absolute Error), were compared and the authors found that the absolute error scores from the active JPR test and the AMEDA test were not significantly correlated. Further, a lack of correlation was also identified between the absolute error scores from the active JPR and the AUC scores from the AMEDA, supporting their conclusion that ‘different cognitive processes underlie the performance of discrimination and reproduction’ (Yang et al., 2020).

When results from these studies are considered together, it can be concluded that none of the methods used for testing proprioception have produced scores that are strongly inter-correlated, and that differential sensitivity scores obtained from the three methods - TTDPM, JPR and AMEDA - that have been employed to test proprioception, in a manner corresponding to the three psychophysical methods originally developed by Fechner (1860) for testing differential sensory acuity, are test-specific (Yang et al, 2018b; Yang et al, 2020). However, the results reported in studies on measuring proprioceptive ability might only hold for proprioception and not for other sense modalities. Stevenson (1918) used pressure on the extended finger as a ‘pressure’ test, and found scores from it not to be highly correlated with visual and auditory differential sensitivity tests, whereas these modalities were highly inter-correlated. Similarly, Deary et al (2004) found that scores from their visual and auditory discrimination tasks were correlated, but that weight discrimination scores did not correlate with either. Thus, there might be something different about proprioception, and it may be that Fechner’s experimental methods are only different for proprioception.

Fechner’s three methods have been widely used in testing visual sensitivity, but the relationship between the scores obtained from the same participants on visual differential sensitivity tasks, such as discriminating between line lengths using different methods, is still unknown. Thus, the purpose of the present study is to determine the relationship between sensitivity scores obtained from the reproduction and discrimination psychophysical methods with visual stimuli.

II. METHODS

Participants. Thirty right-handed undergraduate students - 15 males and 15 females, aged between 19 and 21 (*M* = 20.5 years, *SD* = 0.89) - were recruited by advertisement from a local university. All participants gave their informed consent prior to their inclusion in this study. In a repeated-measures design, participants were tested with three visual differential sensitivity tests designed on the basis of the methods of constant stimuli, adjustment, and absolute judgement. Because one of the tests in this study involved drawing lines on paper, participants’ handedness was assessed by the Chinese version of the Edinburgh Handedness Inventory (Yang et al, 2018a).

Procedures. Each participant undertook the following three tests, one after another, with an interval of 5 minutes between tests. During the test, one experimenter manually showed the testing cards to the participant in the randomised trial sequence, and the other recorded their responses.

Method of Constant Stimuli (MCS). The stimuli were five lines of different length: 2.5cm, 2.75cm, 3cm, 3.25cm, and 3.5cm. The standard stimulus (SS) was 3cms. Two lines would be shown horizontally side by side on a white card, one of which was always the standard of 3cms, and the other one of the five variable stimuli. Each variable stimulus (VS) line was paired with the standard ten times, with the standard randomly on the right 5 times and on the left 5 times. The testing sequence of each pair was randomized. There was a 5cm distance between the two displayed lines. During the test, one of the experimenters showed the cards and asked the participant whether the line on the right was longer or shorter than the one on the left. The participant was asked to make a judgement and answer the experimenter with “the right is longer” or “the right is shorter”. Then the other experimenter recorded the response on each trial in the stimulus table indicating the response (“shorter” / “longer”) and whether it was correct or not (Y / N). During the test, the experimenters did not give the participant any feedback regarding their judgement. The participant was required to respond immediately when the paired lines were shown. Then after the participant responded, they would be shown the next pair of lines. Each participant completed all 40 trials.

In MCS, all the responses of each participant were scored to calculate the number of the times the participant responded with “longer than the SS” at each variable location. For example, if the participant had 8 responses *correct* at the 2.5 cm location, then they had 2 *incorrect* (longer than SS) responses. However, if the participant had 8 responses *correc*t at the 3.5 cm location, they must have perceived this as “greater” than the SS on 8 occasions. Then the length value of the variable stimuli, the number of “greater than the SS” responses at each variable stimulus, and the trials made at each variable location

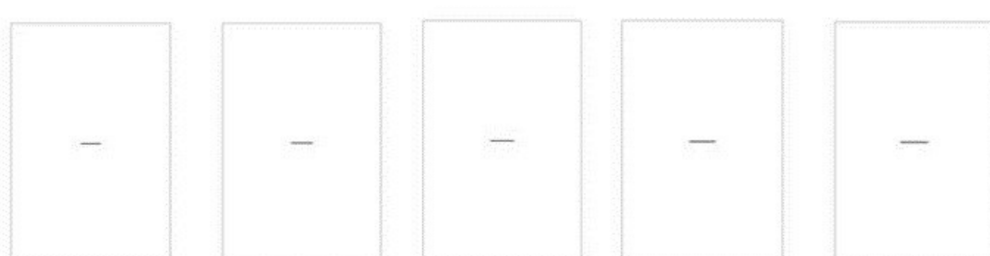


Figure 1. The five horizontal lines that were stimuli in the present study. The lengths of the lines from left to right shown on the white cards were 2.5cm, 2.75cm, 3cm, and 3.5 cm. The width of these five lines was 0.5 cm.

Table 1. Correlations between the Constant Stimuli Test, Adjustment Test, and the Absolute Judgement Test

Test Method	Test Score System		Constant Stimuli Test JND	Adjustment Test AE	Absolute Judgement Test AUC	Absolute Judgement Test AE
Constant Stimuli Test	JND	Pearson Correlation	1	.108	-.362*	.279
		Sig. (2-tailed)		.570	.049	.136
		N	30	30	30	30
Adjustment Test	AE	Pearson Correlation		1	-.079	-.001
		Sig. (2-tailed)			.677	.994
		N		30	30	30
Absolute Judgement Test	AUC	Pearson Correlation			1	-.831**
		Sig. (2-tailed)				.001
		N			30	30
	AE	Pearson Correlation				1
		Sig. (2-tailed)				
		N				30

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

were read into the PROBIT subroutine in SPSS v.23 to calculate the *just noticeably greater* (JNG) point, at a p of 0.75, and the *just noticeably less* (JNL) point at a p of 0.25. Finally, the just noticeable difference (JND) was calculated using the following equation and was used to represent the participant's visual differential sensitivity:

$$JND = \frac{(JNG - SS) + (SS - JNL)}{2}$$

Method of Adjustment/Reproduction (MAR). The Stimuli were the same five lines tested in Test one. Each time only one of the five lines was shown on the card. During the test, the participant was asked to look at the line shown on the card and then reproduce the line on a piece of paper with a pen. When the participant finished reproducing the line on the card, the card was manually switched to the next by an experimenter. During the test, no feedback was given regarding the accuracy of reproduction. Each stimulus line was shown to the participant in the randomized testing sequence for ten times, for 50 trials in total.

In MAR, absolute error (AE) was calculated to represent the participant's visual differential sensitivity. The absolute error (AE), or difference between the actual length of the stimulus line and the line length reproduced by the participant was calculated, then the mean of the AEs for the 50 trials was obtained to represent the participant's visual differential sensitivity in the test.

Method of Absolute Judgement (MAJ). The same stimuli in Test one and Test two were used. The five lines were labeled as No.1 to No.5, the shortest being No.1 and the longest No.5. Before the test, the participant was given a familiarization session, in which the five lines were manually presented on white cards by an experimenter three times, thus 15 times in total, and the experimenter told the participant the corresponding number of each line when shown. During the test, the participant was asked to look at the line shown on the card and make a judgement, telling the experimenter the number they felt corresponded to the line. Each line was manually presented, in a random sequence, 10 times, making 50 times in total. During the test, the experimenter did not give the participant any response regarding their judgement. All responses were recorded by another experimenter.

In MAJ, the raw data of each test were entered into a 5*5 matrix representing the frequency with which each response was made at each length stimulus. Pair-wise Receiver Operating Characteristic (ROC) curves were produced by non-parametric signal detection analysis, namely, by comparing responses to distances 1 and 2, 2 and 3, 3 and 4, 4 and 5 (Macmillan and Creelman, 2005). The ROC curve for each pair of lengths was plotted and the area under the curve (AUC) obtained as a discrimination measure (Maher and Adams 1996; McNicol 2005). Thereafter, the mean of the pair-wise Area Under the Curve (AUC) scores was calculated with SPSS v.23 to obtain a single discrimination score for each participant to represent the participant's visual differential sensitivity. AUC values range from 0.5, representing chance responding, to 1.0, representing perfect discrimination.

Next, the participants' numeric responses in the absolute judgement test were converted to the corresponding stimulus length, in order to produce an AE by comparing this with the length of the actually presented stimulus. The mean of the AEs for absolute judgement tests was calculated to compare with that for adjustment tests and AUC for absolute judgement tests.

The relationship between the results of the three methods – JNDs for Test one, AEs for Test two, and AUC and AEs for Test three - was analyzed in SPSS using Pearson's product-moment correlation. Significance was set at p less than 0.05 throughout.

III. RESULTS

The mean JND for Constant Stimuli Test was 0.791cms (±0.11), the mean AE for Adjustment Test was 0.543cms (±0.29), and the mean AUC and AE for Absolute Judgement Test were 0.788 (±0.55) and 0.136cms (±0.53) respectively. The Adjustment Test AE scores were not significantly correlated with either the Constant Stimuli Test JND or the Absolute Judgement Test AUC, nor to the Absolute Judgement Test AE. The only significant correlation between the different methods was found between the mean JND for the Constant Stimuli Test and the mean AUC for the Absolute Judgement Test, which were negatively correlated with an *r* of – 0.362. The mean AUC and AE obtained from responses in the Absolute Judgement Test were negatively and significantly correlated with each other (r = -0.831).

IV. DISCUSSION

The present study examined the relationship between scores from reproduction and discrimination psychophysical methods, and found that the mean AE from the reproduction test was not significantly correlated with the mean JND from the constant stimuli test, or the mean AUC or mean AE from the absolute judgement test. Only the visual sensitivity scores from the two non-reproduction tests were significantly inter-correlated.

The lack of correlation identified between the AE score from the reproduction test and the AUC and JND scores for visual differential sensitivity is consistent with the findings reported by Yang et al. (2020) in their proprioception study. Regarding the method of reproduction/adjustment (or average error), Matlin (1979) observed that it differs from the other methods because it is the only method where ‘subjects have to make adjustments by themselves’. That is, it is the process of making successive adjustments to the reproduction stimulus until it seems equal to the criterion stimulus that is different to the process of judging presented stimuli that occurs in the Method of Constant Stimuli and in the Absolute Judgment Method. Yang et al. (2020) concluded from their studies that different cognitive processes underlie the performance of discrimination and reproduction tasks. Differences in the memory abilities needed for different proprioception tasks are considered to be the cause of the lack of correlation between methods (Elangovan et al., 2014; Yang et al., 2020). The three major classifications of memory (sensory memory, short-term memory, and long-term memory) represent memory stores of different durations. Sensory memory (SM), lasting only one or two seconds, involves brief storage of sensory information before it is either forgotten or passed on for further processing. Short-term memory (STM) holds small amounts of information for a duration of the order of seconds, and Long-term memory (LTM) stores information for periods of time longer than a few seconds. Han et al. (2016), in their systematic review of the methods for measuring proprioception, proposed that the testing methods of TTDPM, JPR and AMEDA, based on Fechner’s three classical psychophysical methods, require different levels of memory, and that for proprioception, the methods of TTDPM, JPR and AMEDA would largely depend on SM, STM, and LTM, respectively.

For the reproduction task of the visual stimuli in the present study, participants need to remember the target line length only for long enough to reproduce the length of the line immediately. Thus, similar to the reproduction task in proprioceptive JPR test, good performance requires good STM. However, although the method of Constant Stimuli presents the Standard and Comparison Stimulus on each trial, thus apparently making it a STM-dependent task, over several trials, participants appear to ‘learn’ the value of the Standard Stimulus, so it becomes more like a LTM task. Therefore, the lack of correlation that has been consistently reported between reproduction test and discrimination test scores is because they employ different memory stores, with different distributions of ability.

Similarly, the modest correlation between two discrimination tests – method of constant stimuli and absolute judgment – could be expected if they depended on the same underlying memory store. In reviewing their translation of Martin and Muller’s (1899) monograph, and translation of comments on that paper made by Titchener (1905), Murray and Link (2021) noted that Martin and Muller (1899) had introduced the new concept of the ‘absolute impression of the lightness or heaviness of the comparison weight’. Murray and Link (2021) then observe, in parentheses, that Titchener (1905) says ‘the impression can influence the judgment only by way of memory’. Later, Titchener (1905) asks whether the Martin-Muller ‘absoluteness of impression’ might be found in other sense departments beyond lifted weights i.e., ‘in judgments which do not involve active movement’ (p.305). However, the concept of ‘absolute impression of a comparison weight’ raised the possibility that it could be judged on its own, as a single stimulus, without the need of the presentation of a standard stimulus with which to compare it. Woodworth and Schlosberg (1966) state that the Method of Single Stimuli ‘takes advantage of these “absolute impressions” and saves time by omitting the standard altogether’. They attribute the development of The Method of Single Stimuli to Wever and Zener (1928), who define their method as one of ‘absolute judgment upon a single presentation of members of a stimulus series’, where the observer has, through prior presentation of the series of stimuli, ‘a certain degree of knowledge of that series’ (p.492). To the extent that both the Method of Constant Stimuli and the Absolute Judgment/Single Stimuli method involve Martin and Muller’s (1899) ‘absolute impressions’, a positive correlation might be expected between differential sensitivity scores from the two tasks, and that is what was observed here.

V. CONCLUSION

The present study examined the relationship between scores from the same participants performing reproduction and discrimination tests with visual stimuli and found that results from the discrimination methods and the method of adjustment/reproduction were not significantly correlated. The results are consistent with the view that, for both visual and proprioceptive sensory modalities, scores obtained with Fechner’s experimental methods for measuring differential sensitivity applied to the same individuals do not tend to be correlated, and therefore do not preserve the relative ranking of the individuals across the methods. We argue that this is because the methods depend on different cognitive processes underlying the performance of discrimination and reproduction tasks, and that there are different individuals who excel at these different cognitive processes.

References

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684.

Deary, I.J., Bell, P.J., Bell, A.J., Campbell, M.L., Fazal, N.D. (2004). Sensory discrimination and intelligence: Testing Spearman's other hypothesis. *American Journal of Psychology*, *117*, 1–18.

De Jong, A., Kilbreath, S. L., Refshauge, K. M., & Adams, R. (2005). Performance in different proprioceptive tests does not correlate in ankles with recurrent sprain. *Archives of physical medicine and rehabilitation*, 86(11), 2101-2105.

Elangovan, N., Herrmann, A., & Konczak, J. (2014). Assessing proprioceptive function: evaluating joint position matching methods against psychophysical thresholds. *Physical therapy*, 94(4), 553.

Han, J., Anson, J., Waddington, G., & Adams, R. (2013). Proprioceptive performance of bilateral upper and lower limb joints: side-general and site-specific effects. *Experimental Brain Research*, 226(3), 313-323.

Han, J., Waddington, G., Adams, R., Anson, J., & Liu, Y. (2016). Assessing proprioception: a critical review of methods. *Journal of Sport and Health Science*, 5(1), 80-90.

Konczak, J., Krawczewski, K., Tuite, P., & Maschke, M. (2007). The perception of passive motion in Parkinson's disease. *Journal of neurology*, 254(5), 655-663.

Macmillan, N., & Creelman, C. (2005). Detection Theory: A User's Guide. Lawrence Erlbaum Associates. New York, 73.

Maher, C. G., & Adams, R. D. (1996). Stiffness judgments are affected by visual occlusion. *Journal of Manipulative & Physiological Therapeutics,* 19(4), 250.

Martin, L.J. & Muller, G.E. (1899) Zur Analyse der Unterschiedsempfindlichkeit: Experimentelle Beitrage. Leipzig, Germany: J.A.Barth.

Matlin, M.W. (1979) Human Experimental Psychology. Monterey, CA: Brooks-Cole.

McNicol, D. (2005). A primer of signal detection theory. Hove: Psychology Press.

Murray, D.J. & Link, S.W. (2021). The Creation of Scientific Psychology. New York, NY: Routledge.

Nagai, T., Allison, K. F., Schmitz, J. L., Sell, T. C., & Lephart, S. M. (2016). Conscious proprioception assessments in sports medicine: how individuals perform each submodality? *Sports Medicine* (pp. 1-13): SM Online Scientific Resources, Dover, DE.

Silvestri, P., & Fisichelli, V. (1985). What are the psychophysical methods? *Perceptual and motor skills, 61*(3), 683-688.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684.

Titchener, E.B. (1905). Experimental Psychology: A manual of laboratory practice. Volume II. Quantitative experiments. Part II. Instructor's manual. New York, NY: Macmillan.

Waddington, G., & Adams, R. (1999). Discrimination of active plantarflexion and inversion movements after ankle injury. *Australian Journal of Physiotherapy*, 45(1), 7-13.

Westlake, K. P., & Culham, E. G. (2007). Sensory-specific balance training in older adults: effect on proprioceptive reintegration and cognitive demands. *Physical therapy*, 87(10), 1274.

Wever, E. G., & Zener, K. E. (1928).The method of absolute judgment in psychophysics. *Psychological Review*, *35*(6), 466-493.

Woodworth, R.S. & Schlosberg, H. (1954) Experimental Psychology (2nd ed.). New York, NY: Holt, Rinehart & Winston.

Yang, N., Waddington, G., Adams, R., & Han, J. (2018a). 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), 255-273.

Yang, N., Adams, R., Han, J., & Waddington, G. (2018b). Fechner's three experimental methods: do the results correlate? Paper presented at the Fechner Day 2018: the 33rd Annual Meeting of the International Society for Psychophysics, Lüneburg, Germany.

Yang, N., Waddington, G., Adams, R., & Han, J. (2020). Joint position reproduction and joint position discrimination at the ankle are not related. *Somatosensory & motor research*, 37(2), 97-105.

Funding

This work was supported by the National Nature Science Foundation of China (NO.31870936), the National Education Department of China (Humanities and Social Science) (NO. 18YJA890006), the Program of Shanghai Academic Research Leaders (NO.20XD1423200), and the Program of Shanghai Sport Technology (NO. 21C005).

Acoustic Correlates of English Consonant-Vowel-Consonant (CVC) Words Obtained with Multivariate Analysis

Yixin Zhang[1], Yoshitaka Nakajima[2], Kazuo Ueda[3] and Gerard B. Remijn[4]

[1] Human Science International Course, Kyushu University, Fukuoka, Japan
[2] Sound Corporation, Fukuoka, Japan
[3] Department of Human Science, Faculty of Design/Research Center for Applied Perceptual Science/Research and Development Center for Five-Sense Devices, Kyushu University, Fukuoka, Japan
[4] Department of Human Science, Kyushu University/Research Center for Applied Perceptual Science, Kyushu University, Fukuoka, Japan

Abstract— Multivariate analysis methods provide a means to connect the acoustic and phonological features of English speech sounds. Our previous research, in which English obstruents (i.e., plosives, affricates, and fricatives) were subjected to a so-called origin-shifted factor analysis, suggested two factors that could reflect the acoustic natures of English obstruents. One of these two factors, the *low & mid-high* factor, had high factor loadings at two frequencies around 300 Hz and 2300 Hz, respectively. The other factor, the *high* factor, had high loadings around 4100 Hz. To identify the acoustic correlates of other types of English consonants, we added sonorant consonants (i.e., nasals, liquids, and glides) into an analysis of 80 English consonant-vowel-consonant (CVC) words. Another goal of the study was to compare the factor scores between initial consonants and final consonants in the CVC-words. The 80 English CVC-words were split and labeled into pure individual phonemes and analyzed with the origin-shifted factor analysis. All the labeled phonemes were divided into three categories: vowels, sonorant consonants, and obstruents. For these three phoneme categories, the factor scores for the temporal midpoint (50%), as well as for the 25% and 75% time points were identified. Using a different phoneme time point for analysis, however, only had some effect on the distribution of the factor scores of vowels but had not much effect on the factor scores for sonorant consonants and obstruents. In future research, we therefore can use the 50% time point, as in previous studies. Moreover, the present results showed that sonorant consonants were strongly related to the *low & mid-high* factor, while obstruents were strongly related to the *high* factor. This suggests that our analysis could differentiate between the two phoneme types. Finally, initial sonorant consonants occupied the highest positions on the *low & mid-high* factor, while the distribution of initial obstruents was spread out more than final obstruents on the *high* factor. Analyzing phonemes acoustically by multivariate analysis thus can be a fruitful way to understand their phonological nature.

Fechner Day 2021

Schedule

ZOOM: TBA

Symposium 1

(Tuesday, OCT 19, 2021)

08:00-09:45, EST (Americas) 14:00-15:45, CEST (Europe/MidEast) 21:00-22:45, JST (Japan)

Symposium 2

(Wednesday, OCT 20, 2021)

07:00-08:50, EST (Americas) 13:00-14:30, CEST (Europe/MidEast) 20:00-21:30, JST (Japan)

Symposium 3

(Thursday, OCT 21, 2021)

08:00-10:35, EST (Americas) 14:00-16:35, CEST (Europe/MidEast) 21:00-23:35, JST (Japan)

Symposium 4 and Business Meeting

(Friday, OCT 22, 2021)

07:00-09:10, EST (Americas) 13:00-14:10, CEST (Europe/MidEast) 20:00-22:10, JST (Japan)