

Methods and applications in: Perception science

Edited by

Anıl Ufuk Batmaz and Alyssa A. Brewer

Published in

Frontiers in Psychology

Frontiers in Neuroscience



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-3413-7
DOI 10.3389/978-2-8325-3413-7

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Methods and applications in: Perception science

Topic editors

Anil Ufuk Batmaz — Concordia University, Canada

Alyssa A. Brewer — University of California, Irvine, United States

Citation

Batmaz, A. U., Brewer, A. A., eds. (2023). *Methods and applications in: Perception science*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3413-7

Table of contents

- 05 **A Minimally Invasive Method for Observing Wind-Up of Flexion Reflex in Humans: Comparison of Electrical and Magnetic Stimulation**
Tomoya Taniguchi, Tomoaki Alex Kinukawa, Nobuyuki Takeuchi, Shunsuke Sugiyama, Makoto Nishihara, Kimitoshi Nishiwaki and Koji Inui
- 13 **AFENet: Attention Fusion Enhancement Network for Optic Disc Segmentation of Premature Infants**
Yuanyuan Peng, Weifang Zhu, Zhongyue Chen, Fei Shi, Meng Wang, Yi Zhou, Lianyu Wang, Yuhe Shen, Daoman Xiang, Feng Chen and Xinjian Chen
- 25 **The Avatar's Gist: How to Transfer Affective Components From Dynamic Walking to Static Body Postures**
Paolo Presti, Davide Ruzzon, Gaia Maria Galasso, Pietro Avanzini, Fausto Caruana and Giovanni Vecchiato
- 39 **Measuring the impact of suppression on visual acuity in children with amblyopia using a dichoptic visual acuity chart**
Bixia Zhu, Meng Liao and Longqian Liu
- 49 **Assessing Heterogeneity in Students' Visual Judgment: Model-Based Partitioning of Image Rankings**
Miles Tallon, Mark W. Greenlee, Ernst Wagner, Katrin Rakoczy, Wolfgang Wiedermann and Ulrich Frick
- 64 **Use of remote data collection methodology to test for an illusory effect on visually guided cursor movements**
Ryan W. Langridge and Jonathan J. Marotta
- 84 **Corrigendum: Use of remote data collection methodology to test for an illusory effect on visually guided cursor movements**
Ryan W. Langridge and Jonathan J. Marotta
- 87 **Exploring the law of color presentation of double-sided heterochromatic digital printing for textile**
Yiting Duan, Min Zhang, Jiu Zhou, Hua Zhou, Liming Xu and Heng Yang
- 104 **Multi-scale information fusion network with label smoothing strategy for corneal ulcer classification in slit lamp images**
Linquan Lv, Mengle Peng, Xuefeng Wang and Yuanjun Wu
- 115 **Comparing random dot motion in MATLAB vs. Inquisit Millisecond**
Kimia C. Yaghoubi, Sarah Kabbara, Sara Arian, Hadi Kobaissi, Megan A. K. Peters and Aaron R. Seitz
- 122 **Multisensory reading promotion in academic libraries**
Wenyan Yu, Yiping Jiang, Yanqi Wu and Yanxia Cheng

- 133 **Methods for measuring egocentric distance perception in visual modality**
Bo Dong, Airui Chen, Zhengyin Gu, Yuan Sun, Xiuling Zhang and Xiaoming Tian
- 150 **Effect of mask coverage on face identification in Taiwanese men and women**
Yi-Lang Chen, Cheng-Yu Wu, Shih-Cheng Li, Tai-Min Yu and Shu-Ping Yu
- 157 **Gauging response time distributions to examine the effect of facial expression inversion**
David L. Bimler and Galina V. Paramei
- 172 **Real-life relevant face perception is not captured by the N170 but reflected in later potentials: A comparison of 2D and virtual reality stimuli**
Merle Sagehorn, Marike Johnsdorf, Joanna Kisker, Sophia Sylvester, Thomas Gruber and Benjamin Schöne
- 190 **Enhanced lower-limb motor imagery by kinesthetic illusion**
Weizhen Wang, Bin Shi, Dong Wang, Jing Wang and Gang Liu



A Minimally Invasive Method for Observing Wind-Up of Flexion Reflex in Humans: Comparison of Electrical and Magnetic Stimulation

Tomoya Taniguchi^{1*}, Tomoaki Alex Kinukawa¹, Nobuyuki Takeuchi², Shunsuke Sugiyama³, Makoto Nishihara⁴, Kimitoshi Nishiwaki¹ and Koji Inui^{5,6}

¹ Department of Anesthesiology, Nagoya University Graduate School of Medicine, Nagoya, Japan, ² Neuropsychiatric Department, Aichi Medical University, Nagakute, Japan, ³ Department of Psychiatry and Psychotherapy, Gifu University, Gifu, Japan, ⁴ Multidisciplinary Pain Center, Aichi Medical University, Nagakute, Japan, ⁵ Department of Functioning and Disability, Institute for Developmental Research, Aichi Developmental Disability Center, Kasugai, Japan, ⁶ Department of Integrative Physiology, National Institute for Physiological Sciences, Okazaki, Japan

OPEN ACCESS

Edited by:

Anıl Ufuk Batmaz,
Kadir Has University, Turkey

Reviewed by:

Nobuhiro Watanabe,
Tokyo Metropolitan Institute
of Gerontology, Japan
Paul Yoo,
University of Toronto, Canada

*Correspondence:

Tomoya Taniguchi
gucchi44@med.nagoya-u.ac.jp

Specialty section:

This article was submitted to
Perception Science,
a section of the journal
Frontiers in Neuroscience

Received: 16 December 2021

Accepted: 25 January 2022

Published: 23 February 2022

Citation:

Taniguchi T, Kinukawa TA, Takeuchi N, Sugiyama S, Nishihara M, Nishiwaki K and Inui K (2022) A Minimally Invasive Method for Observing Wind-Up of Flexion Reflex in Humans: Comparison of Electrical and Magnetic Stimulation. *Front. Neurosci.* 16:837340. doi: 10.3389/fnins.2022.837340

Wind-up like pain or temporal summation of pain is a phenomenon in which pain sensation is increased in a frequency-dependent manner by applying repeated noxious stimuli of uniform intensity. Temporal summation in humans has been studied by observing the increase in pain or flexion reflex by repetitive electrical or thermal stimulations. Nonetheless, because the measurement is accompanied by severe pain, a minimally invasive method is desirable. Gradual augmentation of flexion reflex and pain induced by repetitive stimulation of the sural nerve was observed using three stimulation methods—namely, bipolar electrical, magnetic, and monopolar electrical stimulation, with 11 healthy male subjects in each group. The effects of frequency, intensity, and number of repetitive stimuli on the increase in the magnitude of flexion reflex and pain rating were compared among the three methods. The reflex was measured using electromyography (EMG) from the short head of the biceps femoris. All three methods produced a frequency- and intensity-dependent progressive increase in reflex and pain; pain scores were significantly lower for magnetic and monopolar stimulations than for bipolar stimulation ($P < 0.05$). The slope of increase in the reflex was steep during the first 4–6 stimuli but became gentler thereafter. In the initial phase, an increase in the reflex during the time before signals of C-fibers arrived at the spinal cord was observed in experiments using high-frequency stimulation, suggesting that wind-up was caused by inputs of A-fibers without the involvement of C-fibers. Magnetic and monopolar stimulations are minimally invasive and useful methods for observing the wind-up of the flexion reflex in humans. Monopolar stimulation is convenient because it does not require special equipment. There is at least a partial mechanism underlying the wind-up of the flexion reflex that does not require C-fibers.

Keywords: central sensitization, flexion reflex, magnetic stimulation, N-methyl-D-aspartate receptor, short-term plasticity, temporal summation, wind-up, withdrawal reflex

INTRODUCTION

Wind-up or temporal summation of pain is a phenomenon in which pain or firing of spinal dorsal horn neurons is increased in a frequency-dependent manner by applying repetitive noxious stimuli (Mendell and Wall, 1965; Herrero et al., 2000). Some of the mechanisms of wind-up share those of central sensitization, and are thought to be involved in hyperalgesia or chronic pain (Li et al., 1999). Therefore, investigation of the wind-up phenomenon would be useful to investigate certain aspects of pathological pain conditions.

In animal studies, wind-up has been observed as facilitated firing of spinal dorsal horn neurons by repetitive electrical stimulation of peripheral C-fibers (Mendell, 1966) or an increase in the electromyogram of the flexion or withdrawal reflex by repetitive electrical stimulation (Price, 1972). In humans, wind-up has been indirectly studied by observing the increase in pain sensation or magnitudes of the RIII component of the flexion reflex by repetitive electrical or heat stimulation (Arendt-Nielsen et al., 1994; Vierck et al., 1997; Guirimand et al., 2000; Terry et al., 2011). The flexion reflex is stably recorded and reproducible (Willer, 1985). Because the short-term plasticity of pain can be easily and clearly observed in humans, wind-up of the flexion reflex is considered a useful method for understanding the pathophysiology of pain. However, it is not widely used because the measurement is accompanied by severe pain, and only a few studies have been conducted using the method. Therefore, less invasive methods are desired.

In this study, we aimed to establish a minimally invasive method for observing wind-up of the flexion reflex in humans and to make it more commonly available. For this purpose, the effects of repetitive stimulation on flexion reflex and pain were compared among three stimulation methods—namely, magnetic stimulation, bipolar electrical stimulation, and monopolar electrical stimulation. Generally, temporal summation of pain in humans is not considered identical to the wind-up of dorsal horn neurons or flexion reflexes in animal experiments. However, to simplify the expression, we used the term “wind-up” in this study to denote progressive increases in the reflex magnitude and pain.

MATERIALS AND METHODS

The study was conducted on 33 healthy male volunteers (age, 21–57 years; mean, 33.1 years). None of the participants had a history of neurological or pain disorders or substance abuse in the last 2 years. They were free of medication at testing. This study was conducted in accordance with the principles embodied in the Declaration of Helsinki and was approved by the Ethics Committee of the National Institute for Physiological Sciences, Okazaki, Japan (approval number: 21A001). Written informed consent forms were obtained from all participants.

Recordings

The participants were randomly assigned to three different groups receiving stimulation with different methods, with 11 participants in each group. All stimuli were applied to the sural

nerve at the level of the lateral malleolus, and flexion reflexes were recorded from the short head of the biceps femoris muscle, ipsilateral to the stimulation, using surface electromyography (EMG). Ag/AgCl disk recording electrodes were attached to the muscle belly as the cathode and to the tendon as the anode with a 5-cm separation (**Figure 1A**). A band ground electrode was placed midway between the stimulation and the recording electrodes. EMG signals were amplified, filtered (10–2 kHz), and stored (–100 to 400 ms) at a sampling rate of 10 kHz using an EMG/evoked potential measuring system (MEB-2300, Nihon Kohden, Tokyo, Japan). The obtained EMG waveforms were full-wave rectified and the average value in the interval from –100 to 0 ms was subtracted as a direct-current offset. The area under the curve (AUC) in the interval of 70–200 ms after the onset of stimulation was calculated as the magnitude of the RIII component of the flexion reflex. Although this procedure excluded the RII component appearing at 40–60 ms, its occurrence was very rare in this study.

Stimulation Methods

The first group received **bipolar electrical stimulation** with a bipolar electrode (NM-420S, Nihon Kohden) with two felt tips, each of 8 mm diameter, separated by 23 mm. The electrode was placed on the skin and sural nerve, with the anode in the distal position (**Figure 1B**). The stimulus was a rectangular 1-ms single pulse.

Stimulation in the second group was *via* transcutaneous **magnetic stimulation** (**Figure 1C**), which was delivered by Magstim Super Rapid device (The Magstim Company, Whitland, United Kingdom). TMS is generally non-invasive. The stimulation output site of the 8-shaped probe was placed on the sural nerve. The duration of the biphasic waveform was less than 1 ms.

The third group was subjected to **monopolar electrical stimulation** (**Figure 1D**). Two Ag/AgCl electrodes with adhesive

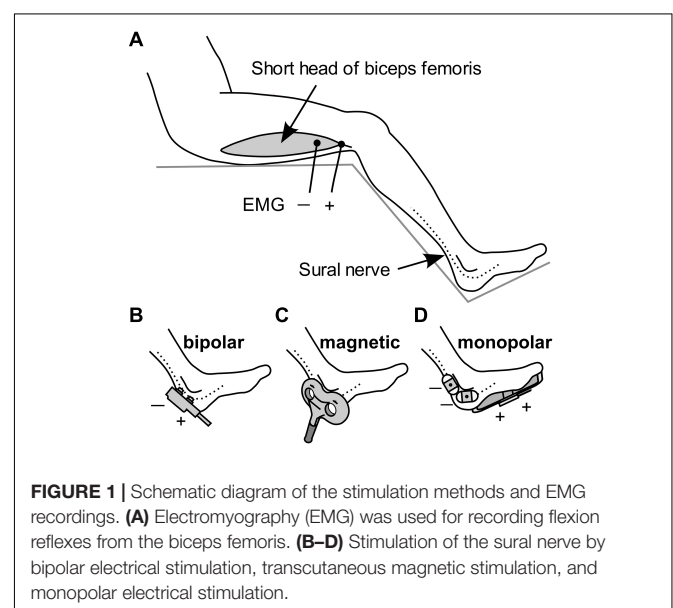


FIGURE 1 | Schematic diagram of the stimulation methods and EMG recordings. (A) Electromyography (EMG) was used for recording flexion reflexes from the biceps femoris. (B–D) Stimulation of the sural nerve by bipolar electrical stimulation, transcutaneous magnetic stimulation, and monopolar electrical stimulation.

conductive gel (Vitrode F 25 mm × 45 mm; Nihon Kohden) were attached to the skin over the sural nerve as the cathode. The anode was a counter electrode with a conductive adhesive gel (180 mm × 115 mm, Valleylab Polyhesive Patient Return Electrode E7507; Covidien, Mansfield, MA, United States) attached to the sole of the foot. The stimulus was a rectangular 1-ms single pulse.

Procedures

In each stimulation group, three experiments were performed with the following procedures: First, the experimental procedures were explained briefly, emphasizing that participants could terminate the experiment at any time if the pain was intolerable. Thereafter, the participants sat with their hip and knee angles at approximately 90° and 130°, respectively.

The pain and reflex thresholds were determined at the beginning. The stimulus was delivered at 0.3 Hz and the intensity was gradually increased. Using an up-and-down procedure, the reflex threshold was determined as the intensity at which the flexion reflex was elicited by 50% of the stimulations. The pain threshold was defined as the intensity at which the tactile sensation turned to pain. Thereafter, three experiments were conducted in random order. There was a 1-minute interval between measurements.

Effects of Stimulus Number (Experiment 1)

The intensity was fixed at the reflex threshold, and 20 consecutive stimuli were applied at 2 Hz. Measurements were taken five times and the average value was calculated. The magnitude of the AUC was plotted against the stimulus number. A segmented linear regression analysis was performed using the segmented package for R (version 4.0.5) (Muggeo, 2017) and the point of the slope change (break point) was determined.

Effects of Stimulation Frequency (Experiment 2)

The stimulation frequency was set at 0.5, 1, 2, 3, 4, or 5 Hz. At each frequency, 10 consecutive reflex threshold stimuli were delivered. The measurement was performed twice with a different order and the mean values were compared.

Effects of Intensity (Experiment 3)

The stimulation intensity was set at 0.5, 0.6, 0.7, 0.8, 0.9, or 1.0 times the reflex threshold. For each stimulation intensity, 10 consecutive stimuli were delivered at 2 Hz. The AUC was compared among the six intensities.

For all measurements, pain scores for the first and last stimuli in a series of consecutive stimuli were recorded using a Numerical Rating Scale. The pain score was defined as follows: 0, no pain at all; and 100, maximum possible pain.

Statistical Analysis

The sample size was calculated to expect the detection of an interaction between stimulation methods and the wind-up effect in pain scores, using G power (version 3.1), with 33 participants (effect size $f = 0.25$, $\alpha = 0.05$, power = 0.8, correlation among repeated measures = 0.65). For each experiment, the AUC values and pain scores were compared among three

methods (stimulation), conditions (intensity or frequency), or stimulus number (wind-up) using two-way or three-way repeated measures analysis of variance (ANOVA) using SPSS 27.0 (IBM Corp, Armonk, NY, United States). For *post-hoc* paired comparisons, a Bonferroni correction was applied. Statistical significance was set at $P < 0.05$.

RESULTS

Flexion Reflex

Some measurements could not be completed because while one participant in the bipolar group experienced intolerable pain, reflex could not be elicited even at the maximal stimulation intensity in two participants in the magnetic group. Therefore, three additional subjects were enrolled so that each stimulation group had 11 subjects. An example of the measurement is shown in Figure 2.

Experiment 1. Effects of Stimulus Number

The averaged waveforms are shown in Supplementary Figure 1. Figure 3 presents the mean magnitudes of reflexes and the results of their analysis by segmented linear regression. The results of two-way ANOVA (stimulation × wind-up) indicated that the stimulation effect was not significant ($F_{2,30} = 1.8$, $P = 0.190$). The wind-up effect of the consecutive stimuli was evident ($F_{19,570} = 37.4$, $P = 2.8 \times 10^{-87}$, partial $\eta^2 = 0.56$). In the segmented regression model, the median (interquartile range) of the breakpoint was 3.3 (2.4–3.9), 6.0 (3.4–7.4), and 5.0 (3.5–8.0) for bipolar, magnetic, and monopolar stimulations, respectively.

Experiment 2. Effects of Stimulation Frequency

The averaged waveforms are shown in Supplementary Figure 2. Figure 4 presents the mean magnitudes of reflexes. Owing to the very short stimulation intervals at 3, 4, and 5 Hz, the recording equipment was unable to pick up the triggers and record the EMG of the even-numbered stimuli; nevertheless, the 10 consecutive stimuli themselves were produced without any problem. Thus, the data with odd stimulus number were used for the analysis. The results of three-way ANOVA (stimulation × frequency × wind-up) showed that the effect of stimulation was not significant ($P = 0.717$), and no interaction was detected. However, wind-up ($F_{4,116} = 120.5$, $P = 2.3 \times 10^{-40}$, partial $\eta^2 = 0.81$) and frequency ($F_{5,145} = 107.6$, $P = 5.4 \times 10^{-47}$, partial $\eta^2 = 0.79$) were significant. The comparisons between frequencies showed significant difference for all pairs, except for the 4–5 Hz pair. The frequency × wind-up interaction was significant ($F_{20,580} = 35.0$, $P = 4.2 \times 10^{-86}$, partial $\eta^2 = 0.55$). When wind-up was tested for each stimulation frequency, all frequencies significantly affected the reflex magnitude.

Experiment 3. Effects of Intensity

The averaged waveforms are shown in Supplementary Figure 3. Figure 5 shows the mean magnitudes. The results of three-way ANOVA (stimulation × intensity × wind-up) showed that intensity ($F_{5,150} = 67.3$, $P = 1.3 \times 10^{-36}$, partial $\eta^2 = 0.69$) and wind-up ($F_{9,270} = 42.6$, $P = 6.0 \times 10^{-47}$, partial $\eta^2 = 0.59$)

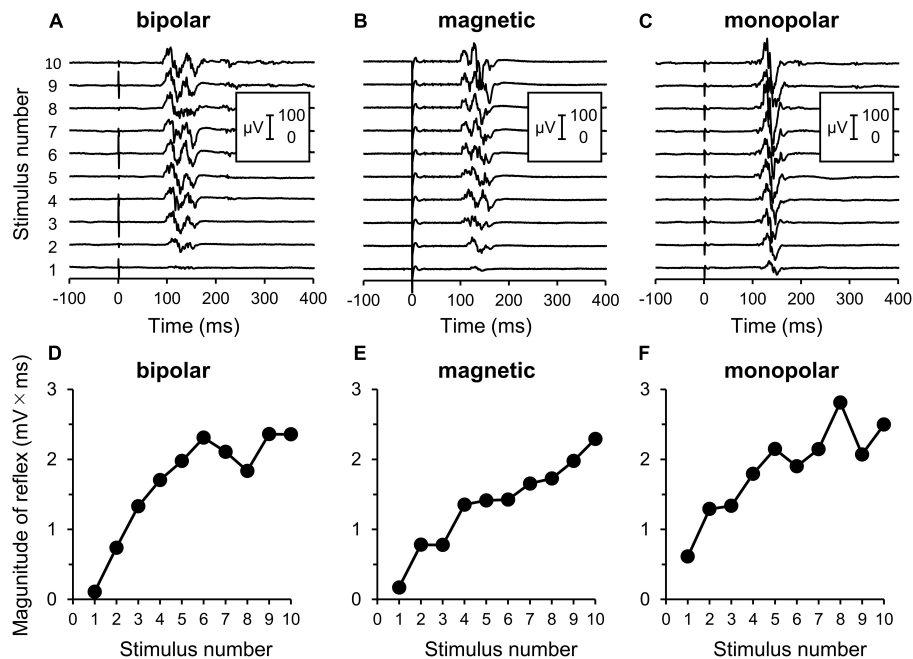


FIGURE 2 | Electromyography (EMG) recordings of flexion reflex and procedures of analyses. (A–C) Original waveforms of a representative subject elicited by a series of 10 consecutive stimuli at the reflex threshold at 2 Hz by bipolar, magnetic, and monopolar stimulations in Experiment 3. (D–F) Relationship between the reflex magnitude [area under the curve (AUC)] and stimulus number. Note that the reflex is clearly enhanced with repeated stimulations.

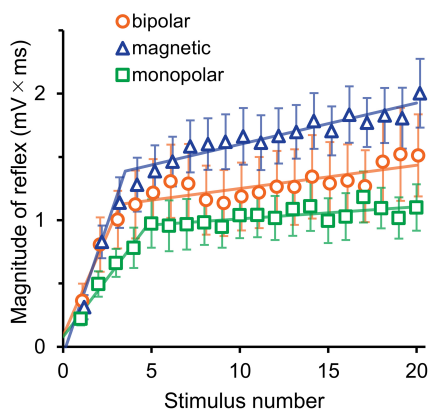


FIGURE 3 | Experiment 1. The mean magnitude \pm SE of the reflex for 20 consecutive stimulations is plotted against the stimulus number. The lines show the results of segmented linear regression analysis.

were significant factors, but stimulation was not ($P = 0.146$). The intensity \times wind-up interaction was significant ($F_{45,1350} = 21.2$, $P = 8.4 \times 10^{-125}$, partial $\eta^2 = 0.41$). *Post-hoc* paired comparisons showed that the magnitude was significantly different among the 10 stimuli only for 0.9 ($P = 6.8 \times 10^{-4}$) and 1.0 ($P = 2.9 \times 10^{-6}$) times the reflex threshold. When the increment of the reflex magnitude between the first and tenth stimuli was compared between the 0.9- and 1.0-times conditions (stimulus \times AUC increment), the wind-up effect was significantly greater for the 1.0-times condition ($P = 1.6 \times 10^{-5}$). The results for the

increments between the first and fifth stimuli were similar ($P = 1.2 \times 10^{-6}$).

Pain Rating

The mean pain score in Experiment 1 is shown in Figure 6A. The score for the last stimulus was significantly greater than that for the first stimulus ($F_{1,30} = 83.0$, $P = 3.8 \times 10^{-10}$, partial $\eta^2 = 0.74$). Although stimulation was a significant factor in determining the pain score ($F_{2,30} = 5.34$, $P = 0.010$, partial $\eta^2 = 0.26$), the stimulation \times wind-up interaction was not ($P = 0.875$), that is, the pain score for bipolar stimulation was significantly greater than that for magnetic ($P = 0.026$) or monopolar ($P = 0.024$) stimulation, but their wind-up effects were not different. There was no significant difference in the pain score between the magnetic and monopolar groups ($P > 0.999$). In Experiment 2, frequency ($F_{5,145} = 96.9$, $P = 1.9 \times 10^{-44}$, partial $\eta^2 = 0.77$), wind-up ($F_{1,29} = 145.7$, $P = 7.8 \times 10^{-13}$, partial $\eta^2 = 0.83$), and stimulation ($F_{2,29} = 10.5$, $P = 3.7 \times 10^{-4}$, partial $\eta^2 = 0.42$) significantly affected the pain score. The overall pain score for bipolar stimulation was also greater than that of magnetic ($P = 5.0 \times 10^{-4}$) and monopolar ($P = 0.004$) ones. *Post-hoc* paired *t*-tests showed that wind-up effects were significant at all stimulation frequencies ($P = 2.2 \times 10^{-14}$ – 8.6×10^{-7}). In Experiment 3, all three factors significantly affected the pain score: intensity ($F_{5,145} = 135.8$, $P = 7.0 \times 10^{-53}$, partial $\eta^2 = 0.82$), wind-up ($F_{1,29} = 67.2$, $P = 4.9 \times 10^{-9}$, partial $\eta^2 = 0.70$), and stimulation ($F_{2,29} = 22.0$, $P = 1.5 \times 10^{-6}$, partial $\eta^2 = 0.60$). The intensity \times wind-up interaction was significant ($F_{5,145} = 29.6$, $P = 1.19 \times 10^{-20}$), and *post-hoc* tests revealed that the pain score

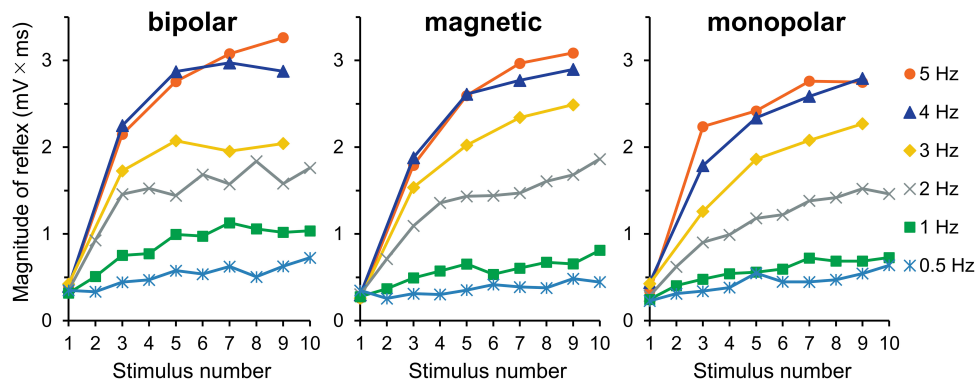


FIGURE 4 | Experiment 2. Mean reflex magnitude for 10 consecutive stimulations at six different stimulation frequencies. Although the wind-up effect is present for all frequency conditions, it is more obvious at higher frequencies.

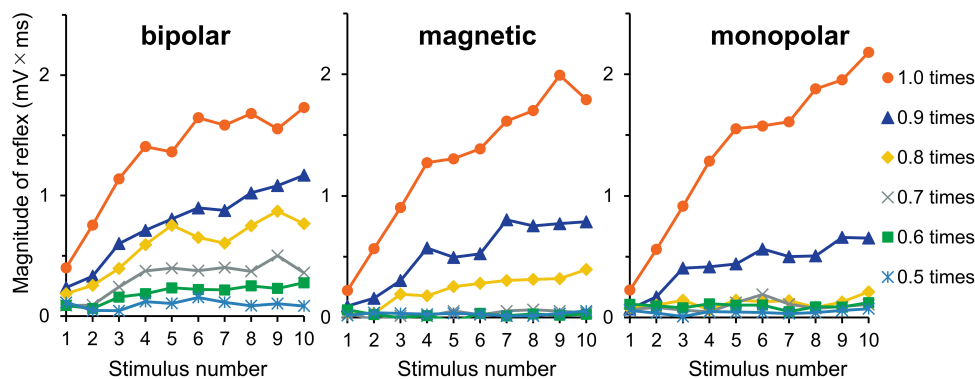


FIGURE 5 | Experiment 3. The reflex magnitude for 10 consecutive stimulations at six stimulation intensities.

was significantly greater for the last stimulus at all intensities from 0.5 to 1.0 times the reflex threshold ($P < 0.003$). Similar to other experiments, the overall pain score for bipolar stimulation was greater than that for magnetic ($P = 1.0 \times 10^{-5}$) and monopolar ($P = 1.0 \times 10^{-5}$) stimulations.

Because the stimulation intensity would have affected the first stimulus pain in Experiment 3, the increment in the pain score between the first and tenth stimuli was compared among the six intensities. As shown in **Figure 6B**, the pain increment gradually increased as the stimulation intensity increased. The results of two-way ANOVA (stimulation \times intensity) showed that intensity ($F_{5,145} = 29.6$, $P = 1.2 \times 10^{-20}$, partial $\eta^2 = 0.51$) significantly affected the pain increment. *Post-hoc* tests showed that the difference was significant for all pairs, except for the pairs with 0.6–0.7, 0.7–0.8, and 0.9–1.0 times the threshold. Therefore, not only the pain score itself but also the degree of pain increment increased with the increase in stimulation intensity.

DISCUSSION

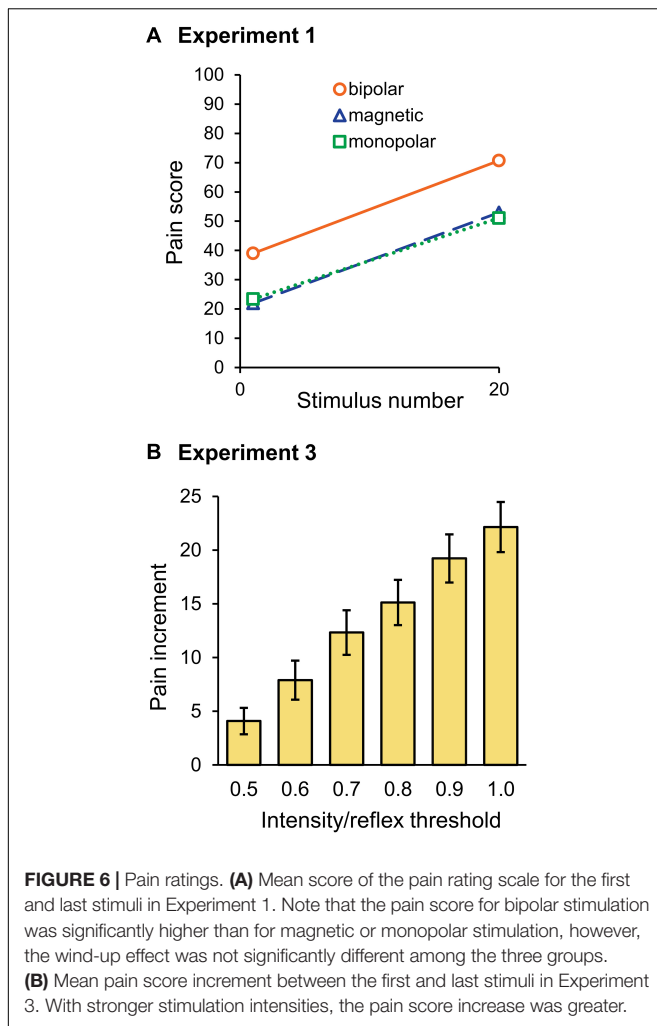
We compared bipolar, magnetic, and monopolar stimulations to establish an appropriate method to observe wind-up. The results indicated that each method exerted a frequency- and

intensity-dependent progressive increase in reflex and pain. Pain sensation was significantly weaker for magnetic and monopolar stimulations than for bipolar stimulation.

Nature of the Response Augmentation

Optimal stimulation frequencies to evoke wind-up phenomenon are approximately 0.5–3 Hz, as was shown for the reflex in humans (Arendt-Nielsen et al., 1994; Terry et al., 2011) and animals (Schouenborg and Sjolund, 1983) and for spinal neurons in animals (Mendell, 1966). Here, wind-up effects were observed for all conditions from 0.5 to 5 Hz. For stimulation intensity, reflex wind-up was observed at 0.9 and 1.0 times the reflex threshold, with greater effect in the 1.0 times the reflex threshold (**Figure 5**) as with a previous study, showing that wind-up is dependent upon stimulation intensity (Arendt-Nielsen et al., 2000). This study's facilitation matches the wind-up concept originally used for spinal nociceptive neurons (Mendell and Wall, 1965).

As shown in **Figure 3**, reflex magnitude increased steeply after repetitive stimulation approximately up to the fifth stimulus, followed by a more gradual increase. This may suggest that the wind-up is mediated by two different mechanisms. As for the initial increase, the wind-up effect was only mediated by



A-fiber input, because the conduction velocity of C-fibers was so slow that signals evoked by the first stimulus did not reach the spinal cord, even when the third reflex occurred by 5 Hz stimulation (Experiment 2). Mechanisms underlying the later phase were uncertain but may include the influence of a ceiling effect, descending inhibitory controls, facilitation by supraspinal mechanisms, C-fiber inputs, and muscle tension. In studies using microneurography, wind-up caused by C-fibers usually shows steep enhancement during initial stimuli, after which the response gradually declines due to a progressive delay in the conduction velocity of C-fibers (Schmelz et al., 1995). Therefore, it is possible that the later phase was due to the gradual decline of the C-fiber activities. However, the present study does not provide direct evidence that C-fiber signals contributed to the wind-up of the reflex.

Overall, pain rating results were comparable to reflex results, consistent with previous studies on temporal pain summation (Vierck et al., 1997; Nielsen and Arendt-Nielsen, 1998; Arendt-Nielsen et al., 2000; Farrell and Gibson, 2007; Terry et al., 2011). Experiment 3 showed that wind-up effect significantly affected pain in all conditions from 0.5 to 1.0 times the reflex threshold,

while the reflex wind-up was significant only for 0.9 and 1.0 times. This is because temporal pain summation is sufficiently induced by pain threshold intensity, which was lower than the reflex threshold. In previous studies using radiant or contact heat stimulation, temporal pain summation was observed for both the first and second pain, with clear dominance for the second pain (Vierck et al., 1997; Nielsen and Arendt-Nielsen, 1998). However, in this study, the pain sensation was not temporally separable in all subjects, suggesting a modest contribution of C-fibers, if any. Therefore, A δ -fibers were probably responsible for pain wind-up.

Methodological Considerations

The flexion reflex in humans has been studied by applying a train of electrical pulses to the sural nerve (Kugelberg et al., 1960; Shahani and Young, 1971; De Willer, 1977; Sandrini et al., 1993, 2005). To the best of our knowledge, no previous study has used other methods to evoke wind-up following sural nerve stimulation. Due to the possibility of a wind-up effect within the train stimulus, a 1-ms single pulse was used in this study. The RIII reflex is known to be suppressed by the occurrence of RII (Hugon, 1973; De Willer, 1977); however, unlike a train, single pulses rarely evoke the RII reflex and make the RIII reflex clearer and facilitate the calculation of the response.

Magnetic stimulation directly activates the nerves with a minimal effect of the skin, thus the stimulation is non-invasive and almost painless (Panizza and Nilsson, 2005; Kizilay et al., 2011). However, in this study, the RIII reflex with obvious pain sensations was caused by magnetic stimulation. Because afferents responsible for the RIII reflex are A δ -fibers (Ertekin et al., 1975), it seemed that A δ -fibers were stimulated by magnetic stimulation. However, due to weaker effects on the skin, the reflex could be observed with lower pain by magnetic stimulation than by bipolar. Some disadvantages of magnetic stimulation include the following: the device is uncommon, its maximum output is relatively weak and sometimes below the reflex threshold, the probe is large, and the device is noisy.

With respect to monopolar stimulation, a weaker pain sensation would be due to the lower current density on the skin caused by the wide electrodes. Unlike magnetic stimulation, monopolar stimulation can be performed with a common electrical neurostimulator. Monopolar stimulation usually has the disadvantage of being less focused than bipolar stimulation (Kombos et al., 1999; Gomez-Tames et al., 2018). However, in this study, it has merits in that the influence of the attachment site is relatively small and the experiment is more reproducible, because the sural nerve generally does not contain motor fibers (Amoiridis et al., 1997) and it is difficult to objectively identify its exact location. In addition, this method could reduce extra pain in the skin caused by stimulating the wrong position displaced from the nerve. Although the tibial nerve afferents in the sole may be stimulated and contribute to the flexion reflex (Ellrich and Treede, 1998), the effect is considered small if present as a large area electrode is used as an anode.

Although the lack of a significant difference of the reflex magnitude among the three stimulation methods may have been due to insufficient detection power with the small sample size, the present results clearly showed that wind-up of the flexion

reflex could be observed with all three stimulation methods and magnetic and monopolar methods resulted in fewer pain scores. Regarding the best stimulation parameters, long measurement times may add uncertain effects, such as facilitation by negative emotions (Fragiotta et al., 2019), suppression by descending inhibitory controls (Gozariu et al., 1997; Ruscheweyh et al., 2011), or alteration of RIIII reflex threshold by central sensitizations (Leone et al., 2021). Stimulation with 5–10 pulses at 4 Hz seems most suitable because wind-up effect was more obvious at higher frequencies, up to 4 Hz. To observe the later slope (Figure 3), stimulation at approximately 2 Hz would be better due to pain tolerance. As 0.9 times the threshold was not sufficient for some participants to observe wind-up, 1.0 times, which was sufficient for all, seems appropriate. Therefore, a practical method for observing the wind-up of the flexion reflex is to use 5–10 consecutive stimuli at 4 Hz or 10–15 stimuli at 2 Hz using single pulses at the reflex threshold with monopolar stimulation.

One important aspect of wind-up is the involvement of N-methyl-D-aspartate (NMDA) receptors in dorsal horn neurons (Herrero et al., 2000; Fossat et al., 2007; Aby et al., 2019) and wind-up of the flexion reflex in humans is indeed inhibited by the NMDA receptor antagonist (Arendt-Nielsen et al., 1995; Guirimand et al., 2000). Therefore, quantification of wind-up in individual subjects would be useful for evaluation of NMDA receptor function and chronic pain diagnosis or psychiatric disorders. Further investigations are needed to clarify how NMDA receptors are involved in the two mechanisms described.

CONCLUSION

The wind-up effect can be observed using all of the three methods. Pain ratings for magnetic and monopolar stimulations were significantly lower than those for bipolar stimulation; however, wind-up effects for flexion reflex and pain did not differ. Although wind-up is thought to be mainly mediated by C-fibers, particularly in animals, there is another mechanism that does not require C-fibers. Observation of wind-up using monopolar stimulation would be useful for investigating short-term plasticity of the nociceptive pathway.

REFERENCES

- Aby, F., Bouali-Benazzouz, R., Landry, M., and Fossat, P. (2019). Windup of nociceptive flexion reflex depends on synaptic and intrinsic properties of dorsal horn neurons in adult rats. *Int. J. Mol. Sci.* 20:6146. doi: 10.3390/ijms20246146
- Amoiridis, G., Schols, L., Ameridis, N., and Przuntek, H. (1997). Motor fibers in the sural nerve of humans. *Neurology* 49, 1725–1728. doi: 10.1212/wnl.49.6.1725
- Arendt-Nielsen, L., Brennum, J., Sindrup, S., and Bak, P. (1994). Electrophysiological and psychophysical quantification of temporal summation in the human nociceptive system. *Eur. J. Appl. Physiol. Occup. Physiol.* 68, 266–273. doi: 10.1007/bf00376776
- Arendt-Nielsen, L., Petersen-Felix, S., Fischer, M., Bak, P., Bjerring, P., and Zbinden, A. M. (1995). The effect of N-methyl-D-aspartate antagonist (ketamine) on single and repeated nociceptive stimuli: a placebo-controlled experimental human study. *Anesth. Analg.* 81, 63–68. doi: 10.1097/0000539-199507000-00013
- Arendt-Nielsen, L., Sonnenborg, F. A., and Andersen, O. K. (2000). Facilitation of the withdrawal reflex by repeated transcutaneous electrical stimulation: an

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the National Institute for Physiological Sciences, Okazaki, Japan (approval number: 21A001). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

TT, SS, and KI: study concept and design, patient recruitment and data collection, data analysis and interpretation, and manuscript preparation. TK, NT, and MN: study concept and design, data collection, data analysis and interpretation, and manuscript preparation. KN: study concept and design, data analysis and interpretation, and manuscript preparation. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Cooperative Study Program (21A001) of the National Institute for Physiological Sciences, Okazaki 444-8585, Japan. Financial support was provided solely by the institutional and departmental sources.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2022.837340/full#supplementary-material>

experimental study on central integration in humans. *Eur. J. Appl. Physiol.* 81, 165–173.

- De Willer, J. C. (1977). Comparative study of perceived pain and nociceptive flexion reflex in man. *Pain* 3, 69–80. doi: 10.1016/0304-3959(77)90036-7

- Ellrich, J., and Treede, R.-D. (1998). Convergence of nociceptive and non-nociceptive inputs onto spinal reflex pathways to the tibialis anterior muscle in humans. *Acta Physiol. Scand.* 163, 391–401. doi: 10.1046/j.1365-201x.1998.t01-1-00392.x

- Ertekin, C., Ertekin, N., and Karcioglu, M. (1975). Conduction velocity along human nociceptive reflex afferent nerve fibres. *J. Neurol. Neurosurg. Psychiatry* 38, 959–965. doi: 10.1136/jnnp.38.10.959

- Farrell, M., and Gibson, S. (2007). Age interacts with stimulus frequency in the temporal summation of pain. *Pain Med.* 8, 514–520. doi: 10.1111/j.1526-4637.2007.00282.x

- Fossat, P., Sibon, I., Le Masson, G., Landry, M., and Nagy, F. (2007). L-type calcium channels and NMDA receptors: a determinant duo for short-term nociceptive plasticity. *Eur. J. Neurosci.* 25, 127–135. doi: 10.1111/j.1460-9568.2006.05256.x

- Fragiotta, G., Pierelli, F., Coppola, G., Conte, C., Perrotta, A., and Serrao, M. (2019). Effect of phobic visual stimulation on spinal nociception. *Physiol. Behav.* 206, 22–27. doi: 10.1016/j.physbeh.2019.03.021
- Gomez-Tames, J., Kutsuna, T., Tamura, M., Muragaki, Y., and Hirata, A. (2018). Intraoperative direct subcortical stimulation: comparison of monopolar and bipolar stimulation. *Phys. Med. Biol.* 63:225013. doi: 10.1088/1361-6560/aaea06
- Gozariu, M., Bragard, D., Willer, J. C., and Le Bars, D. (1997). Temporal summation of C-fiber afferent inputs: competition between facilitatory and inhibitory effects on C-fiber reflex in the rat. *J. Neurophysiol.* 78, 3165–3179. doi: 10.1152/jn.1997.78.6.3165
- Guirimand, F., Dupont, X., Brasseur, L., Chauvin, M., and Bouhassira, D. (2000). The effects of ketamine on the temporal summation (wind-up) of the R(III) nociceptive flexion reflex and pain in humans. *Anesth. Analg.* 90, 408–414. doi: 10.1097/0000539-200002000-00031
- Herrero, J. F., Laird, J. M., and López-García, J. A. (2000). Wind-up of spinal cord neurones and pain sensation: much ado about something? *Prog. Neurobiol.* 61, 169–203. doi: 10.1016/s0301-0082(99)00051-9
- Hugon, M. (1973). Exteroceptive reflexes to stimulation of the sural nerve in normal man. *New Dev. Electromyograp. Clin. Neurophysiol.* 3, 713–729.
- Kizilay, F., Bilgen, R., Aydin Gungor, H., Uysal, H., and Ertekin, C. (2011). Flexor reflexes elicited by magnetic and electric stimulation of the sural nerve. *Neurol. Res.* 33, 609–613. doi: 10.1179/1743132810Y.0000000026
- Kombos, T., Suess, O., Kern, B.-C., Funk, T., Hoell, T., Kopetsch, O., et al. (1999). Comparison between monopolar and bipolar electrical stimulation of the motor cortex. *Acta Neurochirurgica* 141, 1295–1301. doi: 10.1007/s007010050433
- Kugelberg, E., Eklund, K., and Grimby, L. (1960). An electromyographic study of the nociceptive reflexes of the lower limb. Mechanism of the plantar responses. *Brain* 83, 394–410. doi: 10.1093/brain/83.3.394
- Leone, C., Di Lionardo, A., Di Pietro, G., Di Stefano, G., Falco, P., Blockeel, A. J., et al. (2021). How different experimental models of secondary hyperalgesia change the nociceptive flexion reflex. *Clin. Neurophysiol.* 132, 2989–2995. doi: 10.1016/j.clinph.2021.08.018
- Li, J., Simone, D. A., and Larson, A. A. (1999). Windup leads to characteristics of central sensitization. *Pain* 79, 75–82. doi: 10.1016/s0304-3959(98)00154-7
- Mendell, L. M. (1966). Physiological properties of unmyelinated fiber projection to the spinal cord. *Exp. Neurol.* 16, 316–332. doi: 10.1016/0014-4886(66)90068-9
- Mendell, L. M., and Wall, P. D. (1965). Responses of single dorsal cord cells to peripheral cutaneous unmyelinated fibres. *Nature* 206, 97–99. doi: 10.1038/206097a0
- Muggeo, V. M. (2017). Interval estimation for the breakpoint in segmented regression: A smoothed score-based approach. *Austr. NZ J. Stat.* 59, 311–322.
- Nielsen, J., and Arendt-Nielsen, L. (1998). The importance of stimulus configuration for temporal summation of first and second pain to repeated heat stimuli. *Eur. J. Pain* 2, 329–341.
- Panizza, M., and Nilsson, J. (2005). *Activation of peripheral nerve and nerve roots in Magnetic Stimulation in Clinical Neurophysiology*. Amsterdam: Elsevier Inc, 31–41.
- Price, D. D. (1972). Characteristics of second pain and flexion reflexes indicative of prolonged central summation. *Exp. Neurol.* 37, 371–387. doi: 10.1016/0014-4886(72)90081-7
- Ruscheweyh, R., Kreusch, A., Albers, C., Sommer, J., and Marziniak, M. (2011). The effect of distraction strategies on pain perception and the nociceptive flexor reflex (R(III) reflex). *Pain* 152, 2662–2671. doi: 10.1016/j.pain.2011.08.016
- Sandrini, G., Arrigo, A., Bono, G., and Nappi, G. (1993). The nociceptive flexion reflex as a tool for exploring pain control systems in headache and other pain syndromes. *Cephalalgia* 13, 21–27. doi: 10.1046/j.1468-2982.1993.1301021.x
- Sandrini, G., Serrao, M., Rossi, P., Romaniello, A., Cruccu, G., and Willer, J. C. (2005). The lower limb flexion reflex in humans. *Prog. Neurobiol.* 77, 353–395. doi: 10.1016/j.pneurobio.2005.11.003
- Schmelz, M., Forster, C., Schmidt, R., Ringkamp, M., Handwerker, H., and Torebjörk, H. (1995). Delayed responses to electrical stimuli reflect C-fiber responsiveness in human microneurography. *Exp. Brain Res.* 104, 331–336.
- Schouenborg, J., and Sjolund, B. H. (1983). Activity evoked by A- and C-afferent fibers in rat dorsal horn neurons and its relation to a flexion reflex. *J. Neurophysiol.* 50, 1108–1121. doi: 10.1152/jn.1983.50.5.1108
- Shahani, B. T., and Young, R. R. (1971). Human flexor reflexes. *J. Neurol. Neurosurg. Psychiatry* 34, 616–627. doi: 10.1136/jnnp.34.5.616
- Terry, E. L., France, C. R., Bartley, E. J., Delventura, J. L., Kerr, K. L., Vincent, A. L., et al. (2011). Standardizing procedures to study sensitization of human spinal nociceptive processes: comparing parameters for temporal summation of the nociceptive flexion reflex (TS-NFR). *Int. J. Psychophysiol.* 81, 263–274. doi: 10.1016/j.ijpsycho.2011.06.021
- Vierck, C. J. Jr., Cannon, R. L., Fry, G., Maixner, W., and Whitsel, B. L. (1997). Characteristics of temporal summation of second pain sensations elicited by brief contact of glabrous skin by a preheated thermode. *J. Neurophysiol.* 78, 992–1002. doi: 10.1152/jn.1997.78.2.992
- Willer, J. C. (1985). Studies on pain. Effects of morphine on a spinal nociceptive flexion reflex and related pain sensation in man. *Brain Res.* 331, 105–114. doi: 10.1016/0006-8993(85)90719-x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Taniguchi, Kinukawa, Takeuchi, Sugiyama, Nishihara, Nishiwaki and Inui. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



AFENet: Attention Fusion Enhancement Network for Optic Disc Segmentation of Premature Infants

Yuanyuan Peng¹, Weifang Zhu¹, Zhongyue Chen¹, Fei Shi¹, Meng Wang¹, Yi Zhou¹, Lianyu Wang¹, Yuhe Shen¹, Daoman Xiang², Feng Chen^{2*} and Xinjian Chen^{1,3*}

¹ Analysis and Visualization Lab, School of Electronics and Information Engineering and Medical Image Processing, Soochow University, Suzhou, China, ² Guangzhou Women and Children's Medical Center, Guangzhou, China, ³ State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou, China

OPEN ACCESS

Edited by:

Alyssa A. Brewer,
University of California, Irvine,
United States

Reviewed by:

Jiong Wu,
Hunan University of Arts and Science,
China
Yanwu Xu,
Baidu Inc., China

*Correspondence:

Xinjian Chen
xjchen@suda.edu.cn
Feng Chen
eyeguangzhou@126.com

Specialty section:

This article was submitted to
Perception Science,
a section of the journal
Frontiers in Neuroscience

Received: 15 December 2021

Accepted: 09 February 2022

Published: 19 April 2022

Citation:

Peng Y, Zhu W, Chen Z, Shi F,
Wang M, Zhou Y, Wang L, Shen Y,
Xiang D, Chen F and Chen X (2022)
AFENet: Attention Fusion
Enhancement Network for Optic Disc
Segmentation of Premature Infants.
Front. Neurosci. 16:836327.
doi: 10.3389/fnins.2022.836327

Retinopathy of prematurity and ischemic brain injury resulting in periventricular white matter damage are the main causes of visual impairment in premature infants. Accurate optic disc (OD) segmentation has important prognostic significance for the auxiliary diagnosis of the above two diseases of premature infants. Because of the complexity and non-uniform illumination and low contrast between background and the target area of the fundus images, the segmentation of OD for infants is challenging and rarely reported in the literature. In this article, to tackle these problems, we propose a novel attention fusion enhancement network (AFENet) for the accurate segmentation of OD in the fundus images of premature infants by fusing adjacent high-level semantic information and multiscale low-level detailed information from different levels based on encoder-decoder network. Specifically, we first design a dual-scale semantic enhancement (DsSE) module between the encoder and the decoder inspired by self-attention mechanism, which can enhance the semantic contextual information for the decoder by reconstructing skip connection. Then, to reduce the semantic gaps between the high-level and low-level features, a multiscale feature fusion (MsFF) module is developed to fuse multiple features of different levels at the top of encoder by using attention mechanism. Finally, the proposed AFENet was evaluated on the fundus images of preterm infants for OD segmentation, which shows that the proposed two modules are both promising. Based on the baseline (Res34UNet), using DsSE or MsFF module alone can increase Dice similarity coefficients by 1.51 and 1.70%, respectively, whereas the integration of the two modules together can increase 2.11%. Compared with other state-of-the-art segmentation methods, the proposed AFENet achieves a high segmentation performance.

Keywords: optic disc segmentation, multiscale features, attention mechanism, fundus images, premature infants

INTRODUCTION

Retinopathy of prematurity (ROP) and ischemic brain injury resulting in periventricular white matter (PVWM) damage can lead to visual impairment and even blindness of prematurity (Mcloone et al., 2006; Chen and Smith, 2007). In terms of ROP, retinal vascular proliferative blindness disease frequently affects premature infants with low birth weight. It is reported that 53,000 of the 15 million premature infants worldwide require ROP treatment every year (Agrawal et al., 2021). In addition, periventricular leukomalacia or periventricular hemorrhage may cause PVWM damage, which is considered to be a more common cause of visual morbidity than ROP in premature infants (Mcloone et al., 2006).

As the survival rates of preterm infants in modern neonatal intensive care unit continue to improve, the prevalence of neonatal ischemic brain injury and ROP will also increase. Early diagnosis and timely treatment of ROP and PVWM can effectively reduce visual impairment and prevent disease blindness. The diagnosis and treatment of ROP are based on stage, zone, and plus disease, which reflect the severity of ROP (Aaberg, 1987; International Committee for the Classification of Retinopathy of Prematurity, 2005). For the plus disease, the diagnostic procedure for ROP is to estimate the curvature of blood vessels in a predetermined area around the optic disc (OD), while ROP zoning is defined according to the location of the symptom of ROP relative to the OD. Meanwhile, a previous study has shown that the severity of ROP seems to be positively correlated with a higher proportion of vertical to the horizontal optic diameter (Brodsky and Glasier, 1993). In addition, several studies have reported the association between IVH and optic nerve hypoplasia (Burke et al., 1991; King and Cronin, 1993; Algawi et al., 1995; Oberacher-Velten et al., 2006). Therefore, accuracy OD segmentation of prematurity is extremely significant for the auxiliary diagnosis of these two diseases.

The segmentation of OD has always been a research hotspot because of its great significance for detecting other anatomical structures in retinal images. OD often appears as bright red circular or oval areas in fundus images, as shown in **Figure 1**. The irregular OD shape, the diffusion of OD region boundary, and the inconsistency of imaging conditions make OD segmentation very challenging, especially for premature infants. In the past, many related studies on OD segmentation are proposed mainly including traditional algorithms and deep learning (DL) algorithms. Traditional algorithms of OD segmentation use either the intensity of OD region or the point of origination of major vessels for OD localization. The former assumes that the pixel intensity of OD region is higher than that in other parts of the retina (Li and Chutatape, 2001; Walter and Klein, 2001; Chrastek et al., 2002). The main disadvantage of this method is that OD may not be detected correctly in some images because of pathology or uneven illumination. The latter method is based on the assumption that the OD region is the starting

point of the major blood vessels of the eye (Hoover and Goldbaum, 2003; Foracchia et al., 2004; Youssif et al., 2007). However, this method may fail when blood vessels are blocked by lesions. With the development of DL technology in recent years, many convolutional neural network (CNN)-based methods, such as fully convolutional network (FCN) (Long et al., 2015), U-Net (Ronneberger et al., 2015), and their variants based networks, have been developed for OD segmentation without considering any prior knowledge, which can overcome the inabilities of traditional technology. For example, Mohan et al. (2018) introduced a prior CNN named P-Net and cascaded the P-Net with their previously proposed Fine-Net, which can further improve the performance of OD segmentation (Mohan et al., 2019). Al-Bander et al. (2018) used a fully convolutional DenseNet with symmetric U-shaped framework for predicting the OD boundary, which achieved a high segmentation accuracy. Fu et al. (2018) proposed M-Net for joint OD and cup segmentation, which is based on multilabel deep network and polar transformation. Liu et al. (2021) proposed a densely connected depth-wise separable convolutional network (DDSC-Net) for joint optic disc and cup segmentation, which outperforms pOSAL (Kadambi et al., 2020), GL-Net (Jiang et al., 2019), M-Net (Fu et al., 2018), and Stack-U-Net (Sevastopolsky et al., 2019). As the encoder and decoder network (U-Net) was proposed for medical image segmentation, many researchers have focused on modifying the U-Net to further enhance the ability of feature learning. For example, context encoder network (CE-Net) was proposed by Gu et al. (2019) for 2D medical image segmentation, which is based on attention mechanism and outperforms the M-Net in the segmentation of OD. Bhatkalkar et al. (2020) modified the basic architectures of DeepLab v3 + and U-Net models (Ronneberger et al., 2015) by integrating an attention module between the encoder and decoder to obtain the better segmentation of OD. Similarly, many variant networks for semantic segmentation tasks have been proposed and achieved high segmentation performance, such as PSPNet (Zhao et al., 2017), Attention U-Net (Oktay et al., 2018), UNet++ (Zhou et al., 2018), and CPFNet (Feng et al., 2020).

In conclusion, U-Net and its variants based on DL hold promise for automated segmentation of OD in digital fundus images. Recently, Agrawal et al. used the original U-Net for the OD segmentation to assist ROP zoning (Agrawal et al., 2021). As far as we know, this is the first time to study the OD segmentation of preterm infants using DL. Different from the segmentation of OD in adults, the OD segmentation of premature infants is more difficult, mainly for the following two reasons: (1) The fundus images of premature infants often have poor image quality and low contrast due to subjective factors such as illumination and eye movement in the actual shooting process of fundus images; (2) due to preterm birth, the retinal structure of preterm infants is often incompletely developed, resulting in low contrast of fundus images, as shown in **Figure 1**. To handle the above challenges and motivated by the previous successful segmentation networks, we propose a novel attention fusion enhancement network (AFENet) based on the modified encoder and decoder network for automatic segmentation of OD in

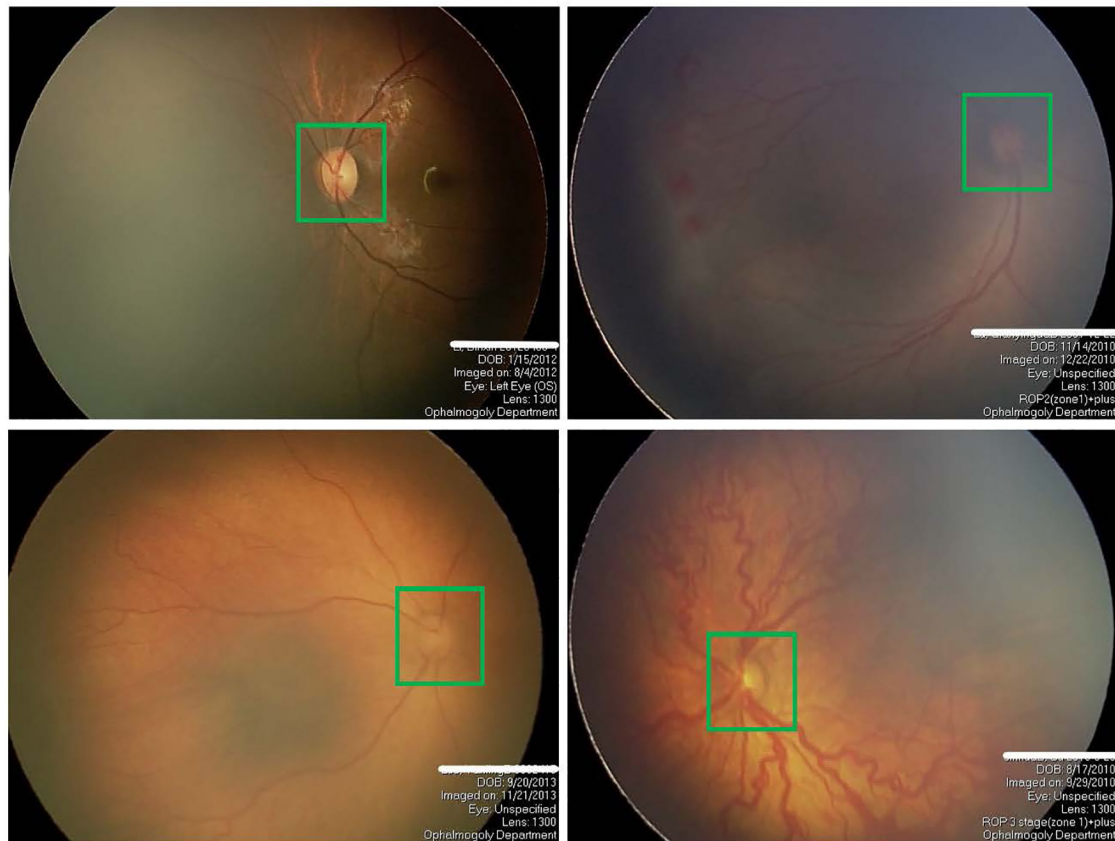


FIGURE 1 | Four examples of fundus images of premature infants, where the optic disc is in the green box.

premature infants. The main contributions of this article can be summarized as follows:

- (1) Two novel attention modules including dual-scale semantic enhancement (DsSE) module and multiscale feature fusion (MsFF) module are developed to fuse adjacent high-level semantic information and multiscale low-level detailed information of different levels between the encoder and decoder, respectively.
- (2) The proposed DsSE module and MsFF module can be easily integrated in U-shape encoder-decoder network and applied for the OD segmentation of premature infants.
- (3) Extensive experiments are conducted to evaluate the effectiveness of the proposed AFENet, and the results show that the proposed AFENet outperforms the state-of-the-art segmentation networks in OD segmentation of premature infants.

The remainder of this article is organized as follows: the proposed method for automatic OD segmentation of premature infants is introduced in section “Methods.” Section “Experiments and Results” presents the experimental results in detail. In section “Conclusion and Discussion,” we conclude this article and suggest future work.

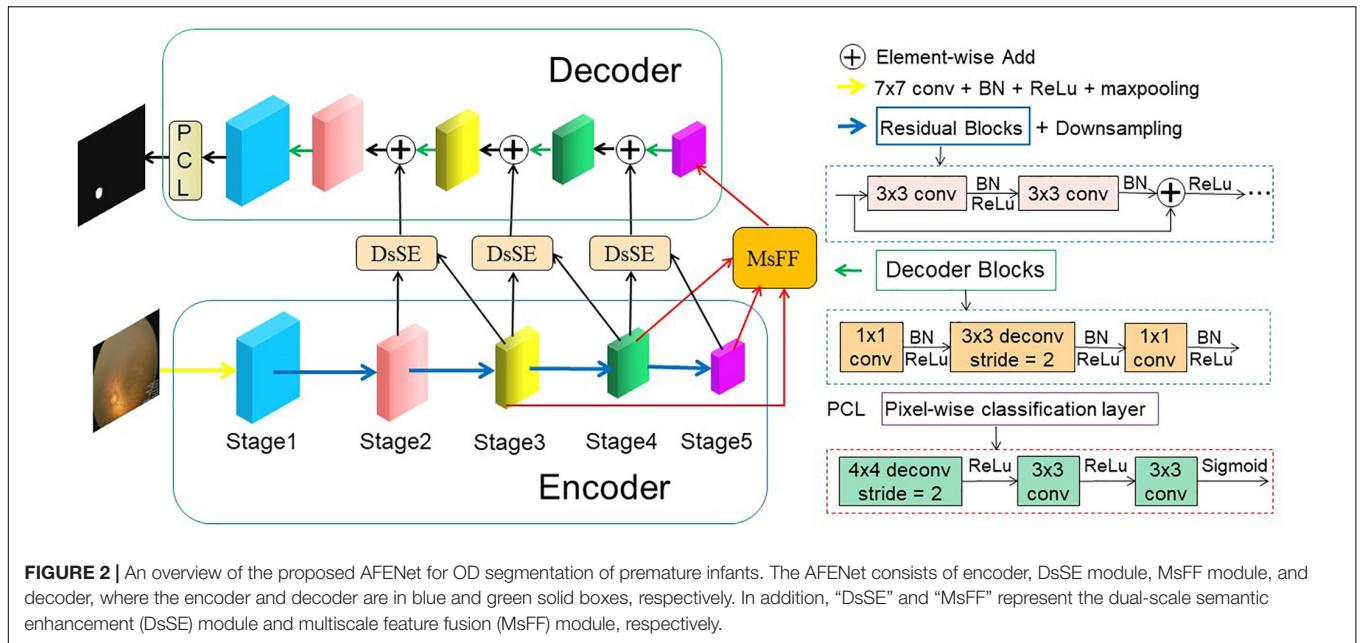
METHODS

Overview

The proposed AFENet for OD segmentation of premature infants is shown in **Figure 2**, which is based on encoder-decoder U-shape architecture and consists of four main parts: feature encoder, DsSE module, MsFF module, and feature decoder. The feature encoder is used to extract spatial features from the input fundus image, whereas the feature decoder is adopted to construct the segmentation map from the encoded features. The DsSE module is embedded between the encoder and decoder to reconstruct the skip connection, whereas the MsFF module is appended on the top of the encoder to fuse the multiscale feature maps from low-level to high-level features, aiming at reducing the semantic gaps between the high-level and low-level features.

Feature Encoder

Different from the original U-Net architecture, where each block of encoder consists of two convolutional layers and a max pooling layer for downsampling, the proposed AFENet uses the pretrained ResNet34 (Apostolopoulos et al., 2017) as the backbone of feature extractor, where the global average pooling layer and the fully connected layer are removed. There are two main reasons to use the pretrained ResNet34 as the backbone



rather than the original U-Net in the encoding part. First, it is inspired by previous studies (Gu et al., 2019; Feng et al., 2020) that the residual blocks with shortcut mechanism in ResNet can accelerate convergence of the network and avoid gradient vanishing, as shown in the right side of **Figure 2**. Second, the experimental results in “Experiments and Results” also show that compared with the original U-Net, the performance of pretrained ResNet34 as backbone in the encoding part had an overall improvement.

Dual-Scale Semantic Enhance Module

In the original U-shape network, the skip connection between the encoder and decoder is concatenation operator, which is used to make up for the loss of fine information caused by downsampling. However, it may ignore global information, introduce irrelevant clusters, and produce semantic gap due to the mismatch of receiving domain (Feng et al., 2020). To solve the above problems and highlight salient features, we designed a DsSE module as shown in **Figure 3**, in which the global semantic information from the next adjacent high-level feature map is fused to enhance the semantic contextual information and reconstruct the skip connection.

In the DsSE module, the skip connection is reconstructed by combining the current feature map with the next adjacent high-level feature map. Suppose that the current input feature map is $F \in \mathbb{R}^{C,H,W}$ and its next adjacent feature map is $N \in \mathbb{R}^{2C,H/2,W/2}$. As can be seen from **Figure 3**, the proposed DsSE module mainly consists of six steps:

- (1) To reduce the dimension of weights and computational cost, a 1×1 convolution is first used to map the feature map N into the same channel space as F , and then we upsample the low-dimension feature map to get the same

size as F via bilinear interpolation, which is denoted as $N_{up} \in \mathbb{R}^{C,H,W}$.

$$N_{up} = \text{Upsample}(\text{Conv1} \times 1(N)) \in \mathbb{R}^{C,H,W} \quad (1)$$

- (2) Three 1×1 convolutions are used to encode the feature map N_{up} to query (Q) and encode the feature map F to key (K) and value (V), respectively.

$$Q = \text{Conv1} \times 1(N_{up}) \in \mathbb{R}^{C/r,H,W} \quad (2)$$

$$K = \text{Conv1} \times 1(F) \in \mathbb{R}^{C/r,H,W} \quad (3)$$

$$V = \text{Conv1} \times 1(F) \in \mathbb{R}^{C,H,W} \quad (4)$$

- (3) We reshape and transpose Q to $Q \in \mathbb{R}^{H*W,C/r}$, and reshape K to $K \in \mathbb{R}^{C/r,H*W}$ and V to $V \in \mathbb{R}^{C,H*W}$, where C , H , and W represent the channel numbers, height, and width of the input feature, and r is the compression ratio and is set to 16 in our study.

$$Q = \text{Transpose}(\text{Reshape}(Q)) \in \mathbb{R}^{H*W,C/r} \quad (5)$$

$$K = \text{Reshape}(K) \in \mathbb{R}^{C/r,H*W} \quad (6)$$

$$V = \text{Reshape}(V) \in \mathbb{R}^{C,H*W} \quad (7)$$

- (4) We calculate the similarity matrix $E \in \mathbb{R}^{H*W,H*W}$ between Q and K to obtain the non-local spatial feature correlation weight guided by global information, as follows:

$$E = \sigma(Q * K) \in \mathbb{R}^{H*W,H*W} \quad (8)$$

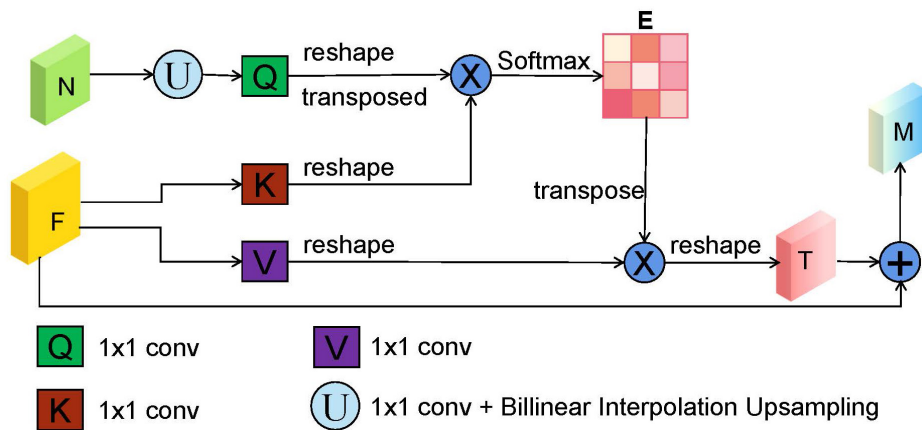


FIGURE 3 | The illustration of dual-scale semantic enhance (DsSE) module. “F,” “N,” “E,” and “M” represent the current input feature map, adjacent next feature map, similarity matrix, and output feature map, respectively. “U” is an upsampling operation, which is obtained by a 1×1 convolution and bilinear interpolation upsampling operation. In addition, “Q,” “K,” and “V” are similar to the three branches of self-attention mechanism (query, key, and value), which are realized by three 1×1 convolutions.

where $*$ is the matrix multiplication operation, and σ is Softmax activation function.

- (5) The similarity matrix E and the corresponding V are weighted by matrix multiplication, and we reshape it to obtain the final spatial response $T \in \mathbb{R}^{C,H,W}$.

$$T = \text{Reshape}(V * E^T) \in \mathbb{R}^{C,H,W} \quad (9)$$

- (6) Finally, we perform element-wise summation operation between T and the current input feature map F to obtain the final attention output $M \in \mathbb{R}^{C,H,W}$ as follows:

$$M = F + T \in \mathbb{R}^{C,H,W} \quad (10)$$

Multiscale Feature Fusion Module

Many previous studies (Oktay et al., 2018; Gu et al., 2019; Feng et al., 2020) have shown that multiscale context information can improve the performance of semantic segmentation, whose core idea is changing global focus to key and local region focus by attention mechanism. However, the above methods may produce the semantic gaps between the low-level and high-level feature maps and ignore the detailed local information. Therefore, to fully utilize the feature interaction between the local context and the global context, an MsFF module is proposed to capture multiscale non-local information with long-range dependency from different levels of encoders, which is illustrated in **Figure 4**.

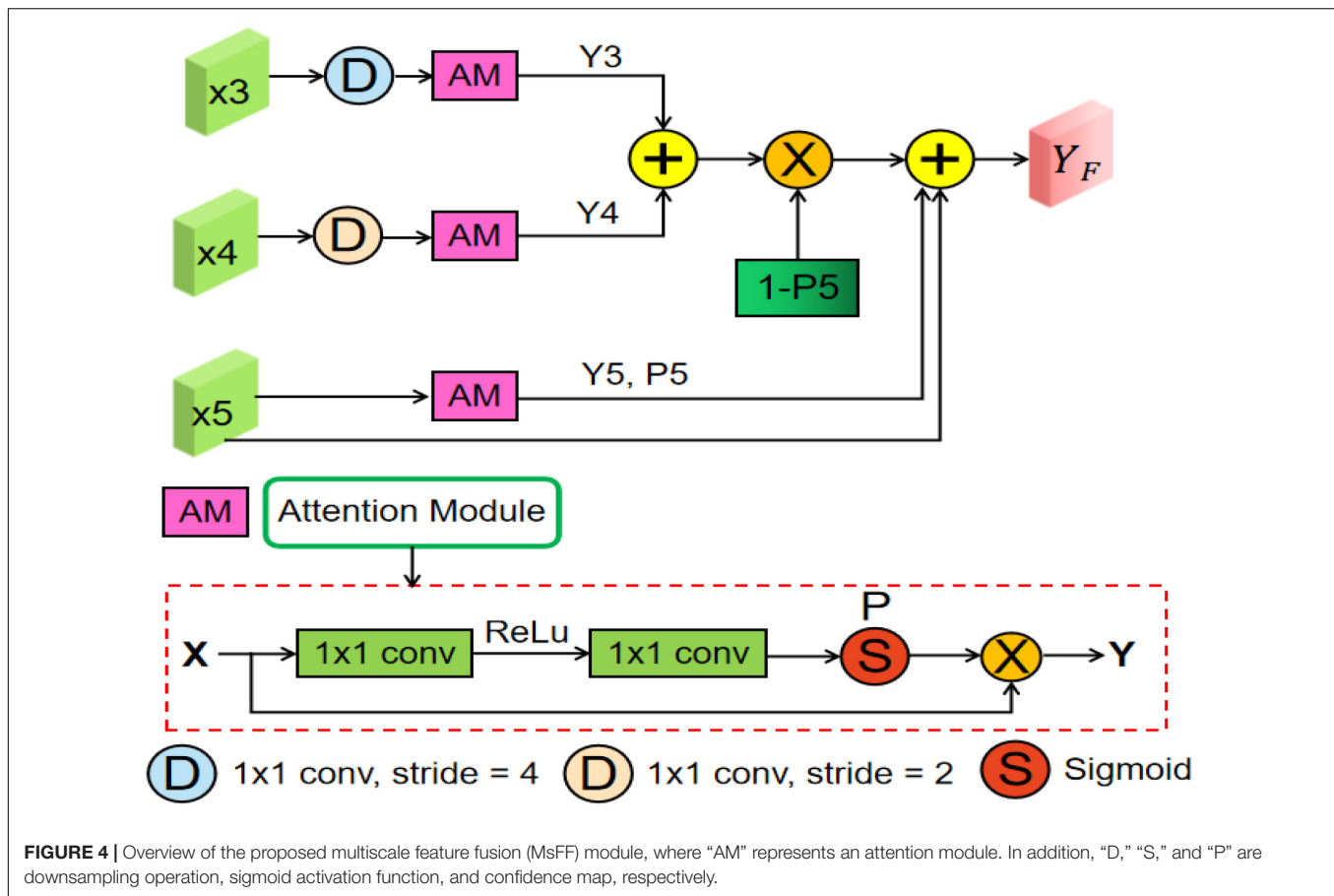
As can be seen from **Figure 2**, the proposed MsFF module is appended on the top of encoder path. Suppose that the feature maps from Stage3, Stage4, and Stage5 are denoted as X_3 , X_4 , and X_5 . As can be observed from **Figure 4**, the feature maps X_3 and X_4 are first downsampled to the same size as X_5 and then feed them and X_5 into the three same attention modules named as AM to generate the corresponding attention feature maps Y_i ($i = 3, 4, 5$) and their corresponding confidence maps P_i ($i = 3, 4, 5$), of which the points with high confidence in P_i ($i = 3, 4, 5$) have a greater possibility to retain the original feature maps values in X_i ($i = 3, 4, 5$), and *vice versa*. As we know, the top

feature map X_5 of encoder has the stronger abstract semantics and the lower spatial resolution without detailed information in segmentation task. Therefore, the confidence map $1 - P_5$ represents the lost detailed information in the top feature map X_5 , where the higher the value of the corresponding position, the richer the detailed information of the corresponding position. In addition, the feature maps Y_3 and Y_4 with limited semantics and rich detailed information are obtained from relatively shallow layers. Thus, based on the above facts, we can supplement the lost detailed information of the corresponding position on the feature map X_5 by using dot product between the confidence map $1 - P_5$ and the sum of attention feature maps Y_3 and Y_4 . Finally, the feature maps with different scales and semantic information are fused to obtain the final feature maps of the top layer with high-level global feature information and low-level local detailed information, as illustrated in Eq. (11). As illustrated in **Figure 4**, the AM consists of two 1×1 convolutional operations, a ReLU activation function, and a sigmoid activation function. Suppose that the input feature of AM is $X \in \mathbb{R}^{C,H,W}$. First, the input feature X is sequentially fed into a convolutional layer with the kernel size of $\frac{C}{r} \times 1 \times 1$, a ReLU activation, a convolutional layer with the kernel size of $C \times 1 \times 1$, and a sigmoid activation function to obtain the confidence map P , where C is channel number, and r is the compression ratio. Then, we multiply P by the input feature X to obtain the final output feature map of AM denoted as $Y \in \mathbb{R}^{C,H,W}$.

$$Y_F = X_5 + Y_5 + (1 - P_5) * (Y_3 + Y_4) \quad (11)$$

Feature Decoder

To restore the high-level semantic feature maps generated by the feature encoder and MsFF module, multiple simple decoder blocks are adopted in the decoder path. Previous studies have shown that the deconvolution could learn a self-adaptive mapping to restore the feature maps with more detailed information (Apostolopoulos et al., 2017; Gu et al., 2019).



Therefore, the deconvolution is adopted in the feature decoder. As can be observed from **Figure 2**, the decoder block mainly consists of a 1×1 convolution, a 3×3 convolution, and a 1×1 convolution consecutively. After the last decoder block, the feature map is restored to $\frac{1}{2}$ of the original input image. Finally, we feed it into a 3×3 deconvolution and two 1×1 convolutions consecutively to obtain the same size segmentation mask as the original input fundus image.

Loss Function

As illustrated in **Figure 2**, our framework is an end-to-end DL network, which takes the fundus images as input and outputs the predicted segmentation results. The proposed AFENet is trained to predict each pixel to be foreground or background, which is a pixel-wise classification problem. A main challenge in medical image segmentation is that the segmentation target (OD) takes a small proportion in the fundus images. To solve the class distribution imbalance problem and similar to Bao et al. (2020), Cheng (2020), Feng et al. (2020), Zhu et al. (2020, 2021), Wang et al. (2021), a joint loss L_{total} is adopted to perform the OD segmentation task, which consists of Dice loss L_{Dice} and binary cross-entropy loss L_{BCE} . The total loss function combined is defined as follows:

$$L_{total} = L_{Dice} + L_{BCE} \quad (12)$$

where,

$$L_{Dice} = 1 - \frac{2|X * Y|}{|X| + |Y|} \quad (13)$$

$$L_{BCE} = - \sum_{h,w} (1 - Y) \log(1 - X) + Y \log(X) \quad (14)$$

where X and Y are the segmentation results and the corresponding ground truth, h and w are the coordinates of the pixel in X and Y . As can be seen from Eqs (13) and (14), the Dice loss and the binary cross-entropy loss are mainly used to optimize the model in the image and pixel levels, respectively.

EXPERIMENTS AND RESULTS

Dataset

In this study, the 1,702 fundus images of premature infants were collected using RetCam3 from Guangzhou Women and Children Medical Center. The gestation ages vary from 26 to 41 weeks, with a mean value of 32 weeks. The collection and analysis of image data were approved by the institutional review board of the Guangzhou Women and Children Medical Center and adhered to the tenets of the Declaration of Helsinki. An information consent was obtained from the guardians of each subject to perform all the imaging procedures. The resolution of

TABLE 1 | Dataset used in this study.

Dataset	Training	Validation	Testing
Num	1,020	341	341

the fundus images was 640×480 . Feng Chen, an ophthalmologist at Guangzhou Women and Children Medical Center, guided the pixel-level annotation. All labeled fundus images were divided into training set, validation set, and testing set, which are shown in **Table 1**.

Experimental Setup

Image Processing

To reduce the computational cost and improve the computational efficiency of the model, all the images were resized to 256×256 by bilinear interpolation and normalized to (0,1). In addition, online data augmentation, including horizontal flipping, rotations of -10 to 10 degrees and affine transformation, was adopted to prevent overfitting and improve the robust ability of the model.

Parameter Setting

The proposed AFE-Net was performed on the public platform Pytorch. We used an NVIDIA RTX3090 GPU with 24-GB memory to train the model with back-propagation algorithm by minimizing the loss function as illustrated in Eq. (12). The Adam was used as the optimizer, where initial learning rate and weight decay were set to 0.0005 and 0.0001, respectively. The batch size and epoch were set to 16 and 100, respectively. To ensure fairness, all the networks in this article were trained with the same optimization schemes, and we saved the best model on validation set in terms of Dice similarity coefficient (Dsc) indicator. The code of the proposed AFENet will be released at <https://github.com/yuanyuanpeng0129/AFENet>.

Evaluation Metrics

To comprehensively and fairly evaluate the segmentation performance of different methods, four evaluation indicators were used, including Dsc and sensitivity (Sen), among which Dsc was the most commonly used metrics in validating the performance of segmentation algorithms (Crum et al., 2006; Milletari et al., 2016; Zhao et al., 2017; Feng et al., 2020). Their definitions are as follows:

$$Dsc = \frac{2 \times TP}{2 \times TP + TN + FP} \quad (15)$$

$$Sen = \frac{TP}{TP + FN} \quad (16)$$

where TP, TN, FP, and FN are true positive, true negative, false positive, and false negative for pixel classification, respectively.

Results

Qualitative Analysis

Figure 5 shows four examples of segmentation results of the proposed AFENet and four classical segmentation networks that are widely used in medical image segmentation tasks, where red represents the correctly segmented region, whereas yellow and blue are the results of false-negative segmentation and false-positive segmentation, respectively. Overall, U-Net performs the worst, especially in the case of blurred OD, which has a serious mis-segmentation problem. There are two possible reasons. First, the simple skip connection of concatenation in the original U-Net ignored global information and may introduce interference from local irrelevant features clutter, which led to the poor performance of U-Net in some medical image segmentation tasks with complex pathological features. Second, multiscale context information, which can consider the structure's surroundings and avoid ambiguous decisions, was not effectively extracted and utilized in each single stage. Compared with U-Net, Att-UNet achieved better segmentation accuracy with few false negatives, which may be due to the introduction of the attention gate (AG) module to guide the model to focus on the salient information in the feature maps of skip connection (**Figures 5B,C**). Similar phenomena also occurred in CE-Net and CPFNet. The performance of CE-Net and CPFNet was better than UNet, which may benefit from the effective combination of pretrained ResNet34 and global/multiscale context information. However, there are still mis-segmentation problems in the segmentation results of Att-UNet, CE-Net, and CPFNet, especially in **Figure 5A**. It is worth noting that the proposed AFENet achieved the best segmentation results with fewer false negatives, especially for the segmentation of fundus images with blurred OD, which are common in fundus images of premature infants.

Quantitative Analysis

We validated the proposed AFENet on the 341 fundus images of premature infants. For convenience, the basic U-shape model with ResNet34 pretrained on ImageNet as the Baseline method. **Table 2** shows the quantitative results of different methods for the OD segmentation. As can be seen from **Table 2**, DeepLabV3 achieved the worst segmentation performance in terms of all segmentation metrics, which is based on FCN architecture and uses dilated convolution to encode multiscale information at the top of encoder. In addition, DeepLabV3 uses only one-step bilinear interpolation for 16-fold upsampling to upsample the feature map of the encoder to the size of the original image, which lead to insufficient detail information, resulting in poor segmentation performance. Although other methods based on FCN architecture achieved better segmentation performances than DeepLabV3, such as FCN (Long et al., 2015), DANet (Fu et al., 2020), GCN (Peng et al., 2017), and PSPNet (Zhao et al., 2017), the problem of detail information loss caused by downsampling still existed, which may cause the poor performance, especially for the fundus images with blurring OD boundary. In addition, the performance of most U-shape-based networks is better than the networks based on FCN architecture, such as U-Net (Ronneberger et al., 2015), Att-UNet

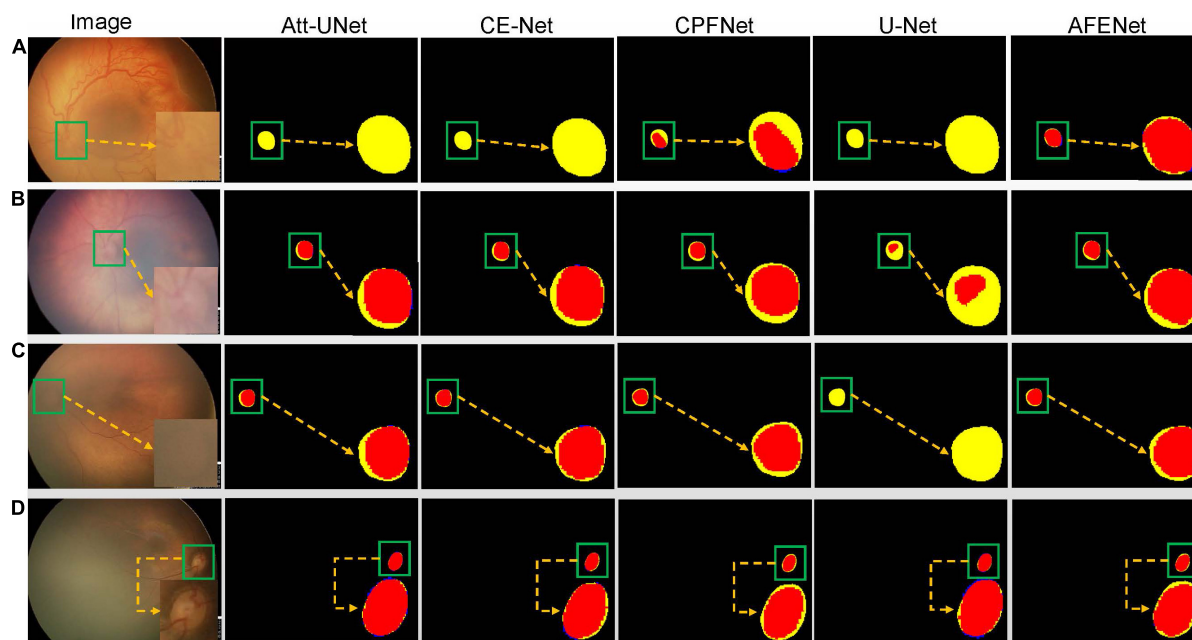


FIGURE 5 | The segmentation results of different methods, where red represents the correctly segmented region, whereas yellow and blue are the results of false-negative segmentation and false-positive segmentation, respectively. To clearly observe the details of the segmentation results, we locally enlarge the segmented target area and place it at the lower right of the corresponding image. In addition, the optic disc is in the green box. **(A)** Represents the fundus image with lesions and blurred optic disc. **(B,C)** Represent normal fundus images with blurred optic disc. **(D)** Represents the normal fundus image with obvious optic disc.

(Oktay et al., 2018), CE-Net (Gu et al., 2019), CPFNet (Feng et al., 2020), CS2Net (Mou et al., 2021), HRSeNet (Wang et al., 2019), DANet (Fu et al., 2020), and TransUNet (Chen et al., 2021), which

TABLE 2 | The results of comparable experiments and ablation studies on OD segmentation of premature infants.

Methods	Dsc (%)	Sen (%)	Parameters (M)
FCN (Long et al., 2015)	89.42	86.95	18.64
DeepLabV3 (Chen et al., 2017)	85.65	90.07	58.16
DANet (Fu et al., 2020)	89.49	91.26	49.48
GCN (Peng et al., 2017)	88.81	89.53	23.62
PSPNet (Zhao et al., 2017)	91.29	91.68	27.76
U-Net (Ronneberger et al., 2015)	89.87	93.56	7.76
Att-UNet (Oktay et al., 2018)	90.91	95.48	8.73
CE-Net (Gu et al., 2019)	91.22	93.74	29.00
CPFNet (Feng et al., 2020)	91.40	89.34	43.27
UNet++ (Zhou et al., 2018)	90.92	92.51	9.16
CS2Net (Mou et al., 2021)	90.64	90.24	8.93
HRSeNet (Wang et al., 2019)	90.88	93.03	1.63
UNet + DsSE + MsFF	0.9130	0.9257	8.09
TransUNet (Chen et al., 2021)	0.9012	0.9314	105.28
Baseline	91.20	91.70	21.66
Baseline + DsSE	92.58	92.35	21.93
Baseline + MsFF	92.75	93.34	21.91
AFENet	93.12	93.22	22.18

Bold values are indicate the best performance.

may be due to the introduction of skip connection between the encoder and the corresponding decoder to alleviate the problem of information loss caused by downsampling. It is worth noting that the proposed AFENet obtains better performance than other segmentation methods in terms of the main evaluation indicator (Dsc). First, compared with Baseline, the performance of the proposed AFENet has been greatly improved, which improves the Dsc and Sen by 2.11 and 1.67% respectively, and achieves 93.12% for Dsc and 93.22% for Sen. Then, compared with other state-of-the-art segmentation networks, the proposed AFENet obtained an overall improvement in terms of the main evaluation indicator Dsc with comparable or less model complexity. For example, compared with the best performance among the comparison methods (CPFNet), the main segmentation evaluation indicator of Dsc of the proposed AFENet increased by 1.88%. Compared with TransUNet (Chen et al., 2021), which has the largest number of model parameters, the performance of the proposed AFENet obtained an overall improvement, especially the Dsc indicator. In addition, compared with GCN (Peng et al., 2017), which has the comparable model complexity, our proposed AFENet has also made great improvement in terms of all evaluation metrics. Especially, we also replaced Backbone with U-Net to further verify the effectiveness and generality of the two modules as shown in Table 2. There are two main findings from Table 2. First, the proposed DsSE module and MsFF module embedded in the U-Net (UNet + DsSE + MsFF) with a small increase in the number of model parameters achieved improvement in terms of the main evaluation indicator Dsc. Second, compared with the UNet + DsSE + MsFF, the proposed AFENet taking the

TABLE 3 | Statistical analysis (*p*-value) of the proposed AFE-Net compared with other CNN-based methods.

Methods	Dsc
AFENet-FCN (Long et al., 2015)	< 1E-4
AFENet-DeepLabV3 (Chen et al., 2017)	< 1E-4
AFENet-DANet (Fu et al., 2020)	< 1E-4
AFENet-GCN (Peng et al., 2017)	< 1E-4
AFENet-PSPNet (Zhao et al., 2017)	0.0006
AFENet-U-Net (Ronneberger et al., 2015)	< 1E-4
AFENet-Att-UNet (Oktay et al., 2018)	0.0004
AFENet-CE-Net (Gu et al., 2019)	0.0016
AFENet-CPFNet (Feng et al., 2020)	< 1E-4
AFENet-UNet++ (Zhou et al., 2018)	0.0055
AFENet-CS2Net (Mou et al., 2021)	0.0001
AFENet-HRSeNet (Wang et al., 2019)	< 1E-4
AFENet-UNet + DsSE + MsFF	0.0004
AFENet-TransUNet (Chen et al., 2017)	< 1E-4
AFENet-Baseline	0.0019

pretrained ResNet34 as Backbone gets an overall improvement in terms of all evaluation indicators (1.99% for Dsc and 0.70% for Sen), which further proves the effectiveness of the pretrained ResNet34 as backbone. These results demonstrate the effectiveness of the proposed AFENet in our task.

Statistical Significance Assessment

To further investigate the statistical significance of the performance improvement by the proposed AFENet over other state-of-art segmentation networks, paired *t*-test was

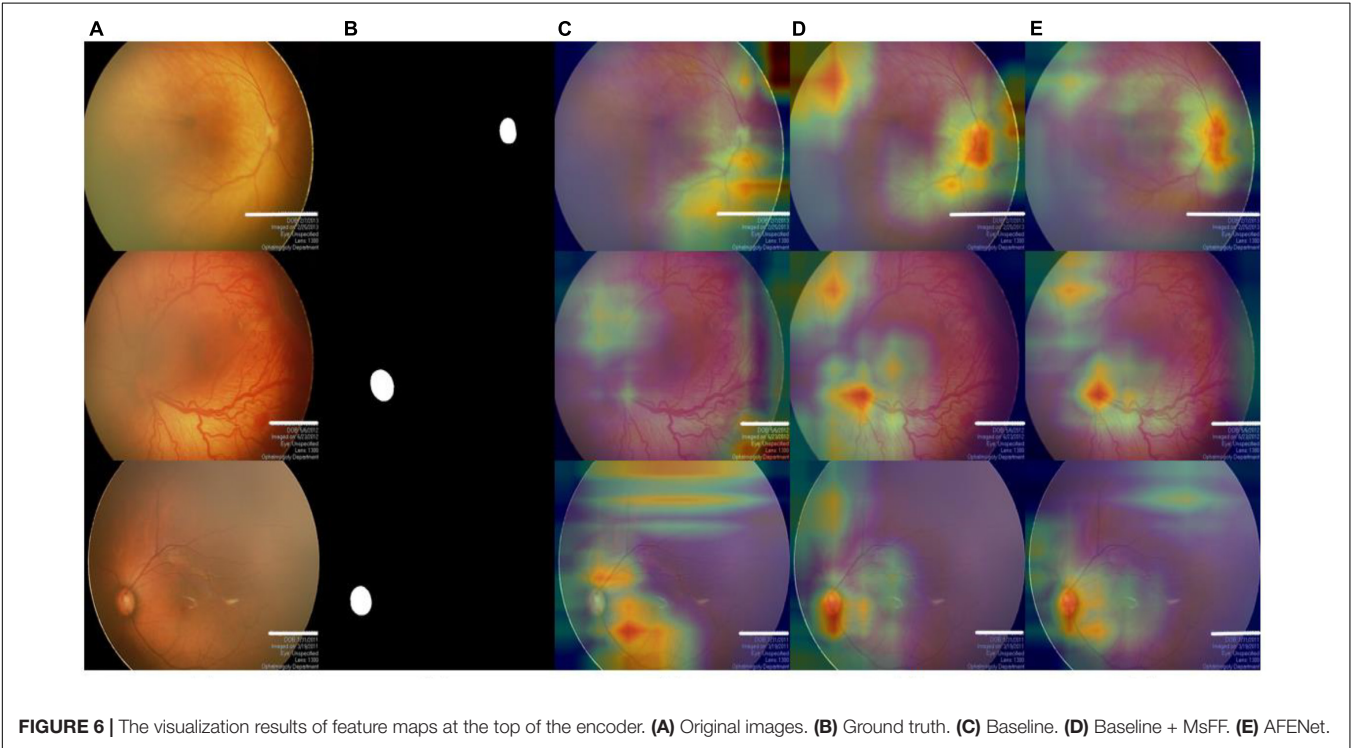
conducted. The *p*-values of the main evaluation indicator (Dsc) are listed in **Table 3**. As can be observed from **Table 3**, all the improvements for Dsc of the proposed AFENet are statistically significant with *p* < 0.05. These results demonstrate that the proposed AFENet can improve the performance of OD segmentation in this study.

Ablation Experiments
Ablation Study for Dual-Scale Semantic Enhancement Module

As can be seen from **Figure 2**, we proposed a novel dual-scale semantic enhancement (DsSE) module to replace the simple skip connection in the original U-shape network. To prove the effectiveness of the proposed DsSE module, we conducted a series of ablation experiments for OD segmentation of premature infants. As shown in **Table 2**, the Baseline + DsSE achieved improvement in terms of all evaluation indexes. Compared with the Baseline, the Dsc and Sen of Baseline + DsSE increased from 91.20 and 91.70% to 92.58 and 92.35%, respectively, which benefits from the fact that the DsSE module can help the U-shape network implicitly learn to suppress irrelevant information and highlight salient features useful for a specific task. These results indicate the effectiveness of the proposed DsSE module.

Ablation Study for Multiscale Feature Fusion Module

It can be observed from **Table 2** that the Baseline + MsFF also obtained an overall improvement in terms of all evaluation indexes. Compared with the Baseline, the Dsc and Sen of the Baseline + MsFF increased by 1.70 and 1.79%, respectively, which benefits the fact that the MsFF module can guide the model to fully utilize the feature interaction between the local



context and the global context and promotes the aggregation of low-level weak semantic information with high-level strong semantic information. In addition, to further demonstrate the effectiveness of the proposed MsFF module, the feature maps are visualized for the qualitative analysis, which are the last convolutional outputs of the encoder and show the focus of the network. We compared the visualization results of MsFF-intergraded network (Baseline + MsFF) with Baseline, as shown in **Figure 6**. It can be seen from **Figures 6A–D** that the proposed MsFF module can focus on the target object regions better than Baseline, which indicates that the proposed MsFF module can effectively aggregate multiscale context information with long-range dependency and improve the segmentation performance.

CONCLUSION AND DISCUSSION

Accurate OD segmentation of premature infants is still a challenging task because of the irregular shapes with various sizes, indistinguishable boundaries, and low contrast between background and OD area in fundus images of premature infants. In this study, to solve these problems, a novel segmentation network named AFENet is proposed to segment the OD in fundus images of premature infants. First, to alleviate the feature learning tendency problem that may be caused by the introduction of simple skip connection between the encoder and the corresponding decoder of the U-shaped based networks, a novel attention module named DsSE module is designed to reconstruct the skip connection, which can enhance the semantic contextual information, highlight the salient features, and improve the ability of model learning. Then, to reduce the semantic gaps between the low-level and high-level feature maps, another novel module named MsFF module was developed to fuse multiple-scale feature maps of different levels by using an attention mechanism, which can fully utilize the feature interaction between the local context and the global context and further improve the segmentation performance. Finally, we conducted a series of experiments on the dataset of fundus images of premature infants to verify the effectiveness of the proposed method. Compared with Baseline, the proposed AFENet with two designed attention modules can adaptively focus on target-related area of fundus images and efficiently improve the segmentation performance of OD, which can be seen from **Figures 6C,E** and **Table 2**. Compared with other state-of-the-art CNN-based segmentation networks, the segmentation performance of the proposed AFENet has been improved significantly, as shown in **Tables 2, 3**. As can be seen from **Figure 5**, compared with the four classical segmentation networks, our AFENet achieves best segmentation results with fewer false negatives and false positives, especially for the segmentation of fundus images with blurred OD, which prove the effectiveness of the proposed AFENet.

The ablation experiments have shown that using the DsSE module or the MsFF module alone can improve the segmentation accuracy, whereas the integration of two modules together can achieve greater improvement. It can be observed from **Table 2**, for Dsc indicator, that the improvement of 1.51, 1.70, and 2.11% can be achieved by using the DsSE module or the MsFF

module alone and the integration of the two modules together, respectively. In addition, taking the proposed MsFF module, for example, the visualization results in **Figures 6C,D** show that the proposed MSFF module can better focus on the location of key area related to OD segmentation than the Baseline, which further demonstrates that the newly designed MsFF module can accurately learn the effective features.

In conclusion, the proposed AFENet holds promise for OD segmentation in fundus images of premature infants and provides the new opportunities and directions for the zoning of ROP and the auxiliary diagnosis of PVWM damage. We believe that our AFENet can also be applied to other medical image segmentation tasks, which requires further exploration and verification. In the future, we will collect more fundus images of premature infants, aiming at focusing on the further performance evaluation of the proposed method and the possibility of diagnosis for other diseases related to premature infants.

DATA AVAILABILITY STATEMENT

The in-house dataset (ROP) presented in this article is not readily available because it is constrained by ethics and patient privacy. Requests to access the datasets should be directed to XC, xjchen@suda.edu.cn.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board of the Guangzhou Women and Children Medical Center and adhered to the tenets of the Declaration of Helsinki. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

YP designed the study, conducted most of experiments, analyzed the experimental results, and drafted the manuscript. WZ and ZC reviewed and revised the manuscript. FS and MW reviewed the manuscript and participated in the design of the experiment. YZ, LW, and YS reviewed the manuscript. DX labeled and interpreted the experimental data in this study. FC collected, labeled and interpreted the experimental data in this study, provided guidance for clinical data analysis, and designed the study. XC designed the study, gave insight in model improvement, reviewed and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported in part by the National Key R&D Program of China (2018YFA0701700) and part by the National Nature Science Foundation of China (U20A20170 and 61622114).

REFERENCES

- Aaberg, T. (1987). An international classification of retinopathy of prematurity: II. The classification of retinal detachment. *Arch. Ophthalmol.* 105, 906–912. doi: 10.1001/archophth.1987.01060070042025
- Agrawal, R., Kulkarni, S., Walambe, R., and Kotecha, K. (2021). Assistive framework for automatic detection of all the zones in retinopathy of prematurity using deep learning. *J. Digit. Imag.* 34, 932–947. doi: 10.1007/s10278-021-00477-8
- Al-Bander, B., Williams, B., Al-Nuaimy, W., Al-Tae, M., Pratt, H., and Zheng, Y. (2018). Dense fully convolutional segmentation of the optic disc and cup in color fundus for glaucoma diagnosis. *Symmetry* 10, 87–102.
- Algawi, K., Beigi, B., Asady, B. M., Murphy, J., and O'keefe, M. (1995). Ophthalmological sequelae following post-haemorrhagic hydrocephalus. *Neuro-ophthalmology* 15, 97–101. doi: 10.3109/01658109509009649
- Apostolopoulos, S., De Zanet, S., Ciller, C., Wolf, S., and Sznitman, R. (2017). "Pathological oct retinal layer segmentation using branch residual u-shape networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (Germany: Springer), 294–301.
- Bao, D., Cheng, X., Zhu, W., Shi, F., and Chen, X. (2020). Attention multi-scale network for pigment epithelial detachment segmentation in OCT images. *Image Process.* 11313, 1131335–1131341. doi: 10.1109/TMI.2022.3142048
- Bhatkalkar, B. J., Reddy, D. R., Prabhu, S., and Bhandary, S. V. (2020). Improving the performance of convolutional neural network for the segmentation of optic disc in fundus images using attention gates and conditional random fields. *IEEE Access* 8, 29299–29310. doi: 10.1109/access.2020.2972318
- Brodsky, M. C., and Glasier, C. M. (1993). Optic nerve hypoplasia: clinical significance of associated central nervous system abnormalities on magnetic resonance imaging. *Arch. Ophthalmol.* 111, 66–74. doi: 10.1001/archophth.1993.01090010070029
- Burke, J. P., O'Keefe, M., and Howell, R. (1991). Optic nerve hypoplasia, encephalopathy, and neurodevelopmental handicap. *Br. J. Ophthalmol.* 75, 236–239. doi: 10.1136/bjo.75.4.236
- Chen, J., Lu, Y., and Lu, Y. (2021). Transunet: Transformers make strong encoders for medical image segmentation. *arXiv arXiv:2102.04306*. [preprint]
- Chen, J., and Smith, L. E. H. (2007). Retinopathy of prematurity. *Angiogenesis* 10, 133–140.
- Chen, L., Papandreou, G., Schroff, F., and Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *arXiv arXiv:1706.05587*. [preprint]
- Cheng, X. (2020). Group-wise attention fusion network for choroid segmentation in OCT images. *Image Process.* 2020, 1131332–1131339.
- Chrastek, R., Wolf, M., Donath, K., Michelson, G., and Niemann, H. (2002). Optic disc segmentation in retinal images. in *Bildverarbeitung für die Medizin* 2002, 263–266.
- Crum, W. R., Camara, O., and Hill, D. L. (2006). Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Transact. Med. Imaging* 25, 1451–1461. doi: 10.1109/TMI.2006.880587
- Feng, S., Zhao, H., Shi, F., Cheng, X., Wang, M., Ma, Y., et al. (2020). CPFNet: Context pyramid fusion network for medical image segmentation. *IEEE Transact. Med. Imaging* 39, 3008–3018. doi: 10.1109/TMI.2020.2983721
- Foracchia, M., Grisan, E., and Ruggeri, A. (2004). Detection of optic disc in retinal images by means of a geometrical model of vessel structure. *IEEE Transact. Med. Imaging* 23, 1189–1195. doi: 10.1109/TMI.2004.829331
- Fu, H., Cheng, J., Xu, Y., Wong, D. W. K., Liu, J., and Cao, X. (2018). Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *IEEE Transact. Med. Imaging* 37, 1597–1605. doi: 10.1109/TMI.2018.2791488
- Fu, J., Liu, J., and Tian, H. (2020). "Dual Attention Network for Scene Segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (New Jersey, NJ: IEEE), 3146–3154. doi: 10.1109/TNNLS.2020.3006524
- Gu, Z., Fu, H., Zhao, Y., Zhou, K., Zhan, T., Hao, H., et al. (2019). CE-Net: Context encoder network for 2D medical image segmentation. *IEEE Transact. Med. Imaging* 38, 2281–2292. doi: 10.1109/TMI.2019.2903562
- Hoover, A., and Goldbaum, M. (2003). Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels. *IEEE Transact. Med. Imaging* 22, 951–958. doi: 10.1109/TMI.2003.815900
- International Committee for the Classification of Retinopathy of Prematurity. (2005). The international classification of retinopathy of prematurity revisited. *Arch. Ophthalmol.* 123, 991–999.
- Jiang, Y., Tan, N., and Peng, T. (2019). Optic disc and cup segmentation based on deep convolutional generative adversarial networks. *IEEE Access* 7, 64483–64493. doi: 10.1109/TMI.2019.2899910
- Kadambi, S., Wang, Z., and Xing, E. (2020). Wgan domain adaptation for the joint optic disc-and-cup segmentation in fundus images. *Int. J. Comput. Assist. Radiol. Surgery* 15, 1205–1213. doi: 10.1007/s11548-020-02144-9
- King, K. M., and Cronin, C. M. (1993). Ocular findings in premature infants with grade IV intraventricular hemorrhage. *J. Pediatr. Ophthalmol. Strabismus* 30, 84–87. doi: 10.3928/0191-3913-19930301-05
- Li, H., and Chutatape, O. (2001). "Automatic location of optic disk in retinal images," in *Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205)*, (New Jersey, NJ: IEEE), 837–840.
- Liu, B., Pan, D., and Song, H. (2021). Joint optic disc and cup segmentation based on densely connected depthwise separable convolution deep network. *BMC Med. Imaging* 21:14. doi: 10.1186/s12880-020-00528-6
- Long, J., Shelhamer, E., and Darrell, T. (2015). "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (New Jersey, NJ: IEEE), 3431–3440.
- McLoone, E., O'Keefe, M., Donoghue, V., McLoone, S., Horgan, N., and Lanigan, B. (2006). RetCam image analysis of optic disc morphology in premature infants and its relation to ischaemic brain injury. *Digest World Core Med. J.* 90, 465–471. doi: 10.1136/bjo.2005.078519
- Milletari, F., Navab, N., and Ahmadi, S.-A. (2016). "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*. New Jersey, NJ: IEEE, 565–571.
- Mohan, D., Harish Kumar, J. R., and Sekhar Seelamantula, C. (2018). "High-Performance optic disc segmentation using convolutional neural networks," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, (New Jersey, NJ: IEEE), 4038–4042. doi: 10.1016/j.bspc.2019.01.022
- Mohan, D., Harish Kumar, J. R., and Sekhar Seelamantula, C. (2019). "Optic Disc Segmentation Using Cascaded Multiresolution Convolutional Neural Networks," in *2019 IEEE International Conference on Image Processing (ICIP)*, (Taipei: IEEE), 834–838.
- Mou, L., Zhao, Y., Fu, H., Liu, Y., Cheng, J., Zheng, Y., et al. (2021). CS2-Net: Deep learning segmentation of curvilinear structures in medical imaging. *Med. Image Anal.* 67, 101874–101892.
- Oberacher-Velten, I. M., Sendtner, P., Gore, F., and Lorenz, B. (2006). Optic disc morphology in relation to stage of retinopathy of prematurity. *Investigat. Ophthalmol. Visual Sci.* 47:13. doi: 10.1136/bjo.2005.085019
- Oktay, O., Lee, M., Mori, K., Glocker, B., Schlemper, J., Heinrich, M. P., et al. (2018). Attention u-net: Learning where to look for the pancreas. *arXiv arXiv:1804.03999*. [preprint]
- Peng, C., Zhang, X., Yu, G., Luo, G., and Sun, J. (2017). "Large Kernel Matters – Improve Semantic Segmentation by Global Convolutional Network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (New Jersey, NJ: IEEE), 4353–4361.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, eds Navab N., Hornegger J., Wells W., Frangi A. (Germany: Springer), 234–241.
- Sevastopolsky, A., Drapak, S., Kiselev, K., Snyder, B. M., Keenan, J. D., and Georgievskaya, A. (2019). Stack-u-net: refinement network for improved optic disc and cup image segmentation. *Med. Imaging* 10949, 1094928–1094930. doi: 10.1186/s12880-020-00528-6
- Walter, T., and Klein, J. C. (2001). "Segmentation of color fundus images of the human retina: Detection of the optic disc and the vascular tree using morphological techniques," in *International Symposium on Medical Data Analysis*, (Berlin: Springer Verlag), 282–287. doi: 10.1016/j.cmpb.2012.06.006
- Wang, J., Sun, K., and Tianheng, C. (2019). Deep High-Resolution Representation Learning for Visual Recognition. *IEEE transactions on pattern analysis and machine intelligence* 43, 3349–3364. doi: 10.1109/TPAMI.2020.2983686
- Wang, T., Zhu, W., Wang, M., Chen, Z., and Chen, X. (2021). "Cu-Segnet: corneal ulcer segmentation network," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, (New Jersey, NJ: IEEE), 1–4.

- Youssif, A. R., Ghalwash, A. Z., and Ghoneim, A. R. (2007). Optic disc detection from normalized digital fundus images by means of a vessels direction matched filter. *IEEE Transact. Med. Imaging* 27, 11–18. doi: 10.1109/TMI.2007.900326
- Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017). “Pyramid scene parsing network,” in *IEEE Computer Vision and Pattern Recognition*, (New Jersey, NJ: IEEE), 6230–6239.
- Zhou, Z., Siddiquee, M., Tajbakhsh, N., and Liang, J. (2018). “UNet++: A nested U-Net architecture for medical image segmentation,” in *4th Deep Learning in Medical Image Analysis (DLIA) Workshop*. (Spain, Cham: Springer), 3–11. doi: 10.1007/978-3-030-00889-5_1
- Zhu, L., Feng, S., Zhu, W., and Chen, X. (2020). ASNet: An adaptive scale network for skin lesion segmentation in dermoscopy images. *Biomed. Appl. Mol. Structural Func. Imaging* 11313, 113108–113114.
- Zhu, Q., Luo, G., Chen, X., Shi, F., Pan, L., Zhu, W., et al. (2021). Joint optic disc and cup segmentation based on multi-module U-shaped network. *Image Process.* 115961, 115960000–115960007.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Peng, Zhu, Chen, Shi, Wang, Zhou, Wang, Shen, Xiang, Chen and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Avatar's Gist: How to Transfer Affective Components From Dynamic Walking to Static Body Postures

Paolo Presti^{1,2*}, Davide Ruzzon^{3,4}, Gaia Maria Galasso², Pietro Avanzini¹, Fausto Caruana¹ and Giovanni Vecchiato^{1*}

¹ Institute of Neuroscience, National Research Council of Italy, Parma, Italy, ² Department of Medicine and Surgery, University of Parma, Parma, Italy, ³ TUNED, Lombardini22, Milan, Italy, ⁴ Dipartimento Culture del Progetto, University IUAV, Venice, Italy

OPEN ACCESS

Edited by:

Anıl Ufuk Batmaz,
Kadir Has University, Turkey

Reviewed by:

Christos Mousas,
Purdue University, United States
Dominik M. Endres,
University of Marburg, Germany
Christian Graff,
Université Grenoble Alpes, France

*Correspondence:

Paolo Presti
paolo.presti@unipr.it
Giovanni Vecchiato
giovanni.vecchiato@in.cnr.it

Specialty section:

This article was submitted to
Perception Science,
a section of the journal
Frontiers in Neuroscience

Received: 23 December 2021

Accepted: 27 April 2022

Published: 15 June 2022

Citation:

Presti P, Ruzzon D, Galasso GM, Avanzini P, Caruana F and Vecchiato G (2022) The Avatar's Gist: How to Transfer Affective Components From Dynamic Walking to Static Body Postures. *Front. Neurosci.* 16:842433. doi: 10.3389/fnins.2022.842433

Dynamic virtual representations of the human being can communicate a broad range of affective states through body movements, thus effectively studying emotion perception. However, the possibility of modeling static body postures preserving affective information is still fundamental in a broad spectrum of experimental settings exploring time-locked cognitive processes. We propose a novel automatic method for creating virtual affective body postures starting from kinematics data. Exploiting body features related to postural cues and movement velocity, we transferred the affective components from dynamic walking to static body postures of male and female virtual avatars. Results of two online experiments showed that participants coherently judged different valence and arousal levels in the avatar's body posture, highlighting the reliability of the proposed methodology. In addition, esthetic and postural cues made women more emotionally expressive than men. Overall, we provided a valid methodology to create affective body postures of virtual avatars, which can be used within different virtual scenarios to understand better the way we perceive the affective state of others.

Keywords: virtual reality, valence, arousal, dynamic walking, body posture

INTRODUCTION

The rapid development of virtual technologies makes it possible to investigate human behavior in fictive environmental and social scenarios, which are otherwise difficult to reproduce and study within standard laboratory settings (Sanchez-Vives and Slater, 2005; Slater and Sanchez-Vives, 2016; Presti et al., 2021). In this context, the dynamic information of body kinematics allows virtual representations of human behavior with specific emotional contents. Researchers showed body kinematics provide significant recognition accuracy of emotions (Atkinson et al., 2004), and postural information allows the discrimination of emotion intensity (Aviezer et al., 2012), demonstrating that body cues are fundamental for a comprehensive understanding of the other's emotion (de Gelder, 2006, 2009; de Gelder et al., 2010). The possibility to model virtual static bodily configurations preserving a sense of dynamicity and affective information is still of particular interest, given the predominant use of static bodily expressions in a broad spectrum of experimental settings that explore time-locked cognitive processes such as event-related potential with electroencephalography and evoked potential with transcranial magnetic stimulation and reaction times in behavioral studies. Moreover, virtual environments without a realistic depiction of human behavior can be uninteresting, resulting in a lack of attention, often

required to study cognitive processes in social scenarios. Creating an expressive virtual character is difficult because of the complex nature of human non-verbal behavior, such as body posture, and there is surprisingly little research on models that generate affective behavior (Coulson, 2004; de Gelder, 2006; Vinayagamoorthy et al., 2008). Therefore, the issue to solve is identifying bodily variables that can be used to build a model of affective behavioral cues.

A strategy to model the posture of a virtual avatar with emotional content involves using motion capture technologies to record an actor's movement and then select the most expressive posture based on subjective judgment (De Silva and Bianchi-Berthouze, 2004; Kleinsmith et al., 2006). Another way is to exploit 3D modeling software to artificially shape the avatar posture by changing the angles between contiguous body parts (Clavel et al., 2009; Buisine et al., 2014). Other systems characterizing emotional expressions depend on the degree of subjective inference and granularity of the measure, whose combination deeply impacts the reliability and efficiency of the categorization in terms of time coding. For instance, the Body Action and Posture Coding System consists of 141 behavioral variables, whose combination describes anatomical articulation, form, and functional level of the movement. A reliability study reports that it took 2,280 min to encode the data set of 6.28 min, thus yielding a coding ratio of 1:363 (Dael et al., 2012). The Facial Action Coding System consists of 44 Action Units that can be coded at different levels of intensity with a coding ratio of 1:100 (Cohn et al., 2007). The Laban Movement Analysis describes how the observed motor action uses components of movement and how each component of movement is related to one another with a coding ratio of 1:30 (Bernardet et al., 2019). As these procedures are highly subjective and time-consuming, there is the need to identify an automatic procedure to transfer the affective information from a body kinematic to a static emotional posture. Hence, this study aims to overcome reliability and time coding issues by automatically extracting body features from kinematic data of walking to transfer the corresponding dynamic affective components to static body postures.

Walking is a natural day-to-day motion that can convey different affective states by combining upper- and lower-limb movements (Montepare et al., 1987; Roether et al., 2009; Karg et al., 2010; Zhao et al., 2019). Previous works have shown that virtual representations of human beings can communicate different affective states through emotional walking (McHugh et al., 2010; Hicheur et al., 2013; Randhavane et al., 2019a, 2021). A recent study of Bhattacharya et al. (2020) found that participants correctly recognized the emotion expressed by virtual avatars according to different gait patterns. Similar results were found in a study where emotional walking was used to animate virtual avatars in different virtual scenarios (e.g., a park, street, and garden) (Randhavane et al., 2019b). Previous studies in kinematic-based movement analysis and affective computing returned that both postural and kinematic features are essential for an accurate description of the individual's affective states (Kleinsmith and Bianchi-Berthouze, 2013; McColl and Nejat, 2014; Stephens-Fripp et al., 2017). In this regard, valence and arousal are typically associated with different body

features and are considered crucial characteristics to describe the human affective experience on continuous and dimensional scales (Lindquist et al., 2012; Kuppens et al., 2013; Kragel and LaBar, 2016). Valence dimension is described by postural cues defining the body's shape during the movement. Hence, joint angles between contiguous body segments and the position assumed by specific body parts are crucial cues for identifying the valence level conveyed by the movement. Head and trunk orientation discriminate between positive and negative valence levels. Walking with a downward leaning of the head/trunk highlights unpleasant affective states as such posture is associated with sadness and anger. On the other hand, an upward orientation identifies joyful walking (Karg et al., 2010; Hicheur et al., 2013; Venture et al., 2014; Crenn et al., 2016; Randhavane et al., 2019b). The volume calculated from the expansion of the body in the three-dimensional space is another postural feature widely exploited for the affective characterization of walking. Instead, a compact posture is associated with sad walking, while an expanded one stands for positive expressions (Crenn et al., 2016; Randhavane et al., 2019b). The dimension of arousal is well-described by kinematic cues considering the quantity of motion of the gesture, which is highly correlated with velocity, acceleration, and jerk of the movement (Karg et al., 2010; Sanghvi et al., 2011; Randhavane et al., 2019b) (Nakagawa et al., 2009). Previous studies reported that walking speed is correlated with the arousal level (Roether et al., 2009; Halovic and Kroos, 2018a; Deligianni et al., 2019). Sad walking was characterized by slow movements, while joy and anger walking, typically considered high arousal emotions, were characterized by fast and rapid movements (Bernhardt and Robinson, 2007; Gross et al., 2012; Barliya et al., 2013; Randhavane et al., 2019b).

This study provides a procedure enabling the automatic creation of emotional body postures by identifying the corresponding most salient time frame from a whole kinematic. For this purpose, emotional walking kinematics were described frame per frame by two distinct body features, namely, body pleasantness (BP) and body dynamicity (BD), which were based on the combination of postural cues and movement velocity of male and female actors.

We hypothesize that the time frames automatically selected with different levels of BP and BD should correspond to coherent valence and arousal levels, separately, reflecting those perceived in the corresponding walking actions.

Results of a first experiment showed that participants coherently assigned valence and arousal scores to avatars' body postures with different levels of BP and BD, and that such scores were strongly correlated with those provided on the correspondent walking actions. Findings also showed that the female avatar was judged as more emotionally expressive than the male one. With the hypothesis that the avatar's physical appearance could contribute to the judgment of valence and arousal, we performed a second online experiment by disrupting the coherence between the gender of the actor and that of the avatar. Hence, the male avatar assumed those body postures derived from female actresses and *vice versa*. This experiment revealed that the combination of both esthetic and

postural characteristics makes women appear more emotionally expressive than men.

Overall, we demonstrated that the proposed methodology successfully transferred affective components from emotional walking to static body postures. The use of BP and BD allowed to select representative emotional frames in a reliable and time-efficient way, leading to the automatic creation of affective body postures. The proposed procedure could be exploited to design stimuli for a broad range of experimental studies investigating time-locked cognitive processes related to the perception of emotional body postures.

MATERIALS AND METHODS

Emotional Kinematics

We considered the EMILYA database to select emotional kinematics for this study (Fourati and Pelachaud, 2014, 2018). Such a database includes daily actions performed by actors and recorded through inertial motion capture technology. We considered the 912 simple walking actions comprising the whole body's movement of five female and five male actors. A.mat file (MATLAB, The MathWorks, Inc., Natick, MA, United States) contains the time-varying root-related positions of 28 body joints for each action. They refer to a 3D space where the xy, yz, and xz planes describe the coronal, sagittal, and transverse planes, respectively, and the origin of the axes corresponds to the middle point between the right and left hip.

Body Pleasantness

The valence dimension is generally linked to postural features more than kinematic ones (Nakagawa et al., 2009). Previous findings show that bowing and expansiveness of the body discriminate the pleasantness of the performed action (Karg et al., 2010; Fourati and Pelachaud, 2015, 2018). Here, we defined BP as the feature adopted to transfer the valence information of the walking to a static body posture. Such score was computed as a combination of three distinct body features, namely, the leaning of the head (LH), the position of the head (PH), and the openness of the body (OB), according to the following procedure. For each kinematic, the LH was computed as the time-average distance between the body joints representing the head and the neck in the sagittal direction:

$$\overline{LH} = \frac{\sum_{t=1}^F (p_{t,z\text{ head}} - p_{t,z\text{ neck}})}{F} \quad (1)$$

where $p_{t,z\text{ head}}$ indicates the z coordinate at time t of the body joint related to the head, and $p_{t,z\text{ neck}}$ indicates the z coordinate at time t of the body joint related to the neck, and F stands for the total number of recorded frames for that kinematic.

The PH was computed as the time-average distance of the body joint representing the head and the origin on the z-axis:

$$\overline{PH} = \frac{\sum_{t=1}^F (p_{t,z\text{ head}})}{F} \quad (2)$$

Finally, the openness of the body was defined as the time-average body spatial extension in the transverse, sagittal, and coronal plane and computed as follows:

$$\overline{OB} = \frac{\sum_{t=1}^F [(\max p_{t,x} - \min p_{t,x}) * (\max p_{t,y} - \min p_{t,y}) * (\max p_{t,z} - \min p_{t,z})]}{F} \quad (3)$$

where $\max p_{t,x} - \min p_{t,x}$ stands for the maximum extension of the body in the lateral direction, $\max p_{t,y} - \min p_{t,y}$ for the maximum extension in the vertical direction, and $\max p_{t,z} - \min p_{t,z}$ stands for the maximum extension of the body in the sagittal direction, each one considered at time t.

To describe the valence dimension of the kinematic with a unique value combining the information conveyed by the abovementioned body features, we performed a principal component analysis (PCA) (Abdi and Williams, 2010) using the LH, PH, and OB variables as inputs. We extracted the first principal component explaining the 83% of data variance and defined the BP as the weighted sum of the three input variables. Hence, for each kinematic, we were able to compute the BP frame by frame according to the following formula:

$$BP_t = 0,77 * LH_t + 0,63 * PH_t + 0,10 * OB_t \quad (4)$$

and then select the frame which BP score was the closest to the mean BP of the whole kinematic, as representative of the kinematic:

$$BP = \min_{t \rightarrow F} \left(\left| BP_t - \frac{\sum_{t=1}^F BP_t}{F} \right| \right) \quad (5)$$

Figure 1 shows the body-joint configurations representing the body features adopted to compute the BP, i.e., the LH, PH, and OB, and illustrates the distribution of the 912 walking actions in the plane defined by the first two components of the PCA, according to their value of LH, PH, and OB.

Body Dynamicity

The dynamicity of the performed action typically contains affective information concerning the arousal dimension (Paterson et al., 2001; Pollick et al., 2001; Nakagawa et al., 2009; Dael et al., 2013). To transfer the dynamicity of the whole kinematic to a static body posture, we defined a BD score and computed this metric for each time frame of the kinematic. We identified the frame associated with the maximum BD value and then selected the corresponding body posture as representative of the arousal level of that action. In detail, for each body joint, we first computed the distance between the position at the time t and t + 1 divided by the duration of the time frame:

$$BD_{j,t} = \frac{P_{t+1} - P_t}{T_f} \quad (6)$$

where j and t identify the specific body joint and the time frame. To obtain the frame-by-frame BD of the full-body kinematics, we averaged the BD related to the body joints:

$$\overline{BD}_t = \frac{\sum_{j=1}^N BD_{j,t}}{N} \quad (7)$$

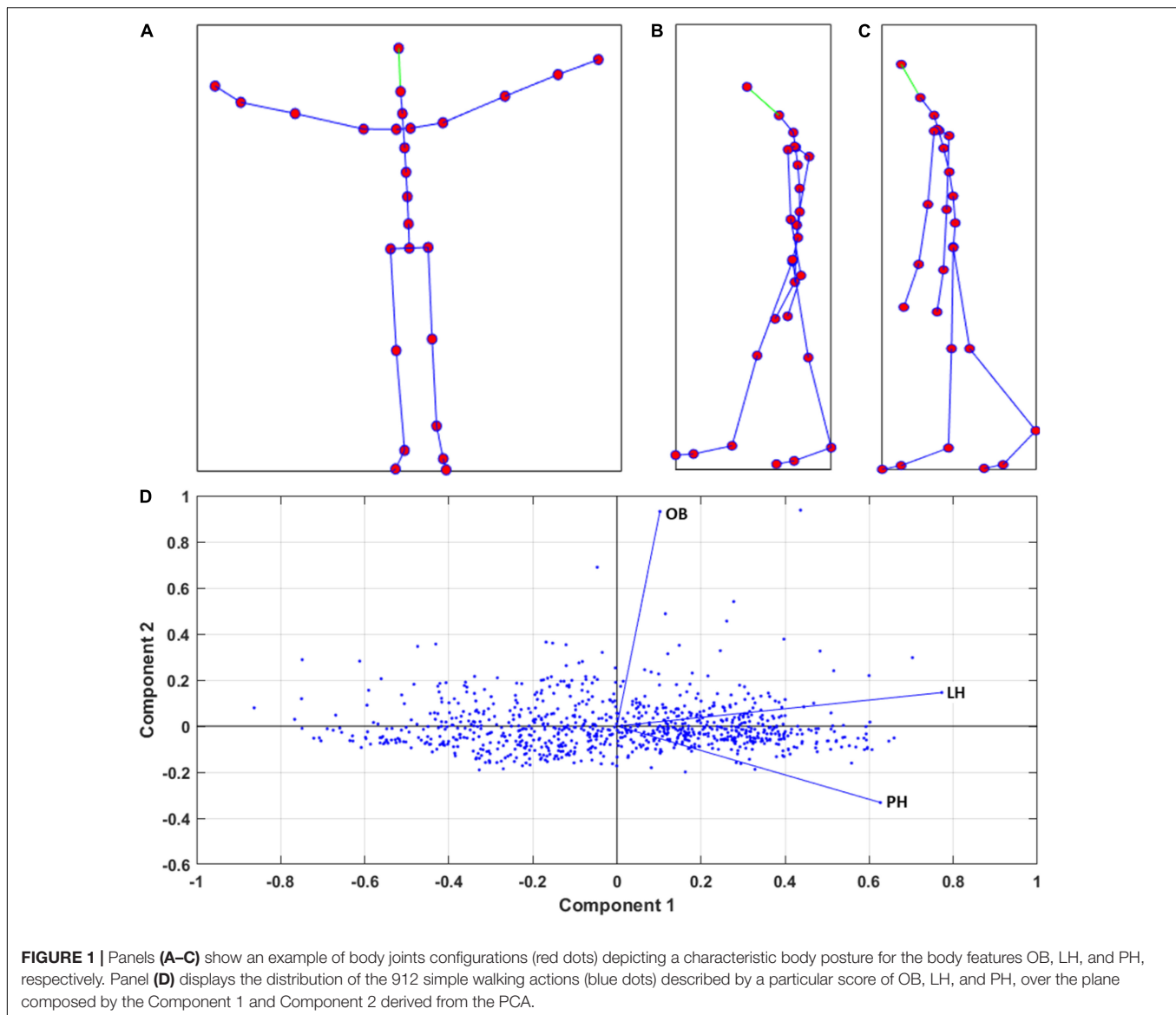


FIGURE 1 | Panels (A–C) show an example of body joints configurations (red dots) depicting a characteristic body posture for the body features OB, LH, and PH, respectively. Panel (D) displays the distribution of the 912 simple walking actions (blue dots) described by a particular score of OB, LH, and PH, over the plane composed by the Component 1 and Component 2 derived from the PCA.

where N is the number of body joints. Finally, we considered the highest BD value across the time frames as representative of the kinematic:

$$BD = \max_{t \rightarrow F} \overline{BD}_t \quad (8)$$

where F stands for the total number of recorded frames for that specific kinematic. **Figure 2** shows body joint configurations representing two body postures with low and high levels of BD. Also, the BD time course of the original walking actions is presented.

Emotional Body Postures

Motion capture data were first converted from .bvh to .fbx file extension using 3ds Max 2020 and following the recommendation described by the Xsens Company.¹ As

¹<https://www.xsens.com/integrations/3dstudiomax>

explained in the previous sections, for each kinematic, we extracted the two configurations of body joints—representative of body postures—associated with the values of expressed arousal and valence. We then selected 90 body postures in three groups corresponding to the low, middle, and high level of BP, each one counting 15 male and 15 female body postures. Analogously, we selected 90 body postures with low, middle, and high values of BD. We performed two one-way ANOVAs with factors BD [$F_{(2,87)} = 1197.5, p < 0.001$] and BP [$F_{(2,87)} = 1187.4, p < 0.001$] to assess differences among the levels low, middle, and high of the corresponding emotional dimension. Bonferroni-corrected pairwise comparisons highlighted significant differences among all the three levels of BD and BP. **Supplementary Figures 1, 2** illustrate the selection procedure of body postures with different levels of BP and BD, respectively. The corresponding statistical analysis shows that BP and BD levels are balanced between male and female actors. Then, we used these data to animate

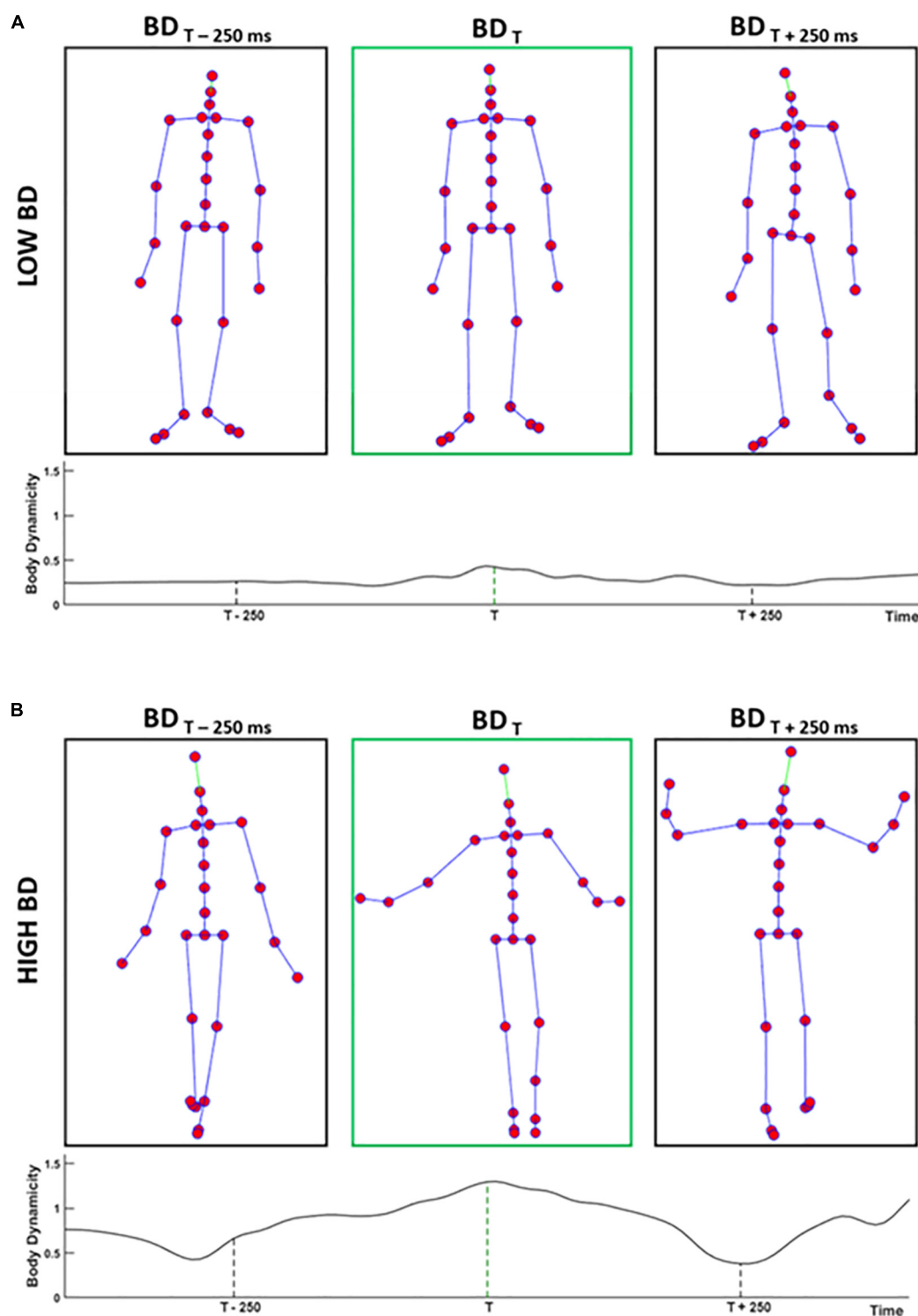


FIGURE 2 | Panels (A,B) show three body joints configurations depicting body postures with low (high) level of BD extracted from two different walkings. The BD time course of the walking is reported below each panel, which shows the precise timepoint from which postures are extracted. Middle panels in green illustrate the representative frames of the whole kinematic, extracted according to the procedure described in the Section “Materials and Methods.” Instead, the first and third panels of the two rows represent two frames extracted 250 ms before and 250 ms after the time of the representative BD, respectively.

two humanoid avatars (one male and one female)² exploiting Unity (2019.1.0f2). Here, we covered the avatar's faces with a skin-colored mask, allowing participants' responses to depend only on the avatar's body posture and not on facial expressions.

EXPERIMENT 1

Experiment 1 consisted of two separate online surveys. In Experiment 1.A, participants rated the valence and arousal levels expressed by static body postures. In Experiment 1.B, participants rated the valence and arousal level conveyed by the corresponding walking actions from which the postures were originally extracted.

Experiment 1.A

Stimuli

The 180 body postures (90 BP, 90 BD) extracted from actions recorded on male (female) actors were assigned to a male (female) avatar to guarantee coherence between the gender of the actor and that of the virtual avatar. **Figure 3** illustrates the representative avatar's body postures with different levels of BP and BD.

Participants, Experimental Procedure, and Data Analysis

We used the online platform Prolific³ to recruit participants for the experiment (Palan and Schitter, 2018). To ensure a reliable sample of participants, we recruited only those who reached almost 95% of the approval rate in previous online experiments and declared to speak English fluently. In line with Prolific policy, participants received a payoff (3.20 £) after completing the experiment. A total of 54 age-matched participants [27 women aged 26.3 ± 5.9 years, 27 men aged 23.4 ± 4.3 years; two-sample t -test, $t_{(52)} = -1.366$, $p = 0.177$] were recruited for the online survey. The sample size was determined using a power analysis computed through the G*Power software (Faul et al., 2007) considering the "as in SPSS" option and setting the significance level (α) at 0.05, the desired power ($1-\beta$) to 0.95, the number of groups to 2, the number of repetition to 6, and the non-sphericity correction ϵ to 1. As we did not find any previous work related to emotional body perception performed online reporting the values of η_p^2 , we set the η_p^2 to 0.16, based on this research performed in a laboratory setting (Kret and de Gelder, 2010), and then doubled the estimate sample size, to compensate the limited control on subject attention and accuracy in doing the online experiment.

Our experiment ran on the online platform Pavlovio.⁴ At the beginning of the experiment, participants read written instructions explaining the concepts of valence and arousal and information on how to express their judgments. Valence was described as the pleasantness state expressed by the body posture of the avatar, referring to the positive or negative character of the event that the body is experiencing. Unpleasant states were

associated with bad feelings or a negative state of mind, while pleasant states were associated with good feelings or a positive state of mind (Colombetti, 2005). The arousal dimension was described as the state of activation expressed by the body posture of the avatar, representing a change of the individual physical and psychological asset. A deactivated state was associated with a low heartbeat, sweating decrease, slow breathing, absence of energy, and decreased attentional and decisional capability. Instead, an activated state was associated with a high heartbeat, sweating increase, fast breathing, feelings of vigor, energy, tension, and increasing attentional and decisional capability (Kreibig, 2010). In each experimental trial, participants judged the arousal and valence level conveyed by the avatar's body posture. Specifically, on the left side of the screen, a picture representing an avatar with a specific body posture was presented, while two questions appeared on the right side with which participants could rate the arousal and valence level expressed by the avatar's body posture. Specifically, as concerns the arousal, they answered the question: "This person looks in a... state" by means of a visual analog scale (VAS) where the lowest value was "deactivated," numerically associated to 0, and the highest one was "activated," numerically associated to 1. As concerns valence, they answered the question: "This person looks in a... state" by means of a VAS where the lowest value was "unpleasant," numerically associated to 0, and the highest one was "pleasant," numerically associated to 1. Participants gave their judgment by clicking the mouse left button on each rating scale and then pressed the space bar to move to the next trial, thus they had no time limits to answer. The whole experiment comprised 180 trials randomly presented in five separate blocks of 36 trials each. Blocks were separated by a self-paced pause during which participants could rest.

Valence and arousal ratings were normalized between 0 and 1 with the *normalize.m* MATLAB function (method, "range"). This function computes a z-score transformation rescaling changes the distance between the minimum and maximum values in a data set by stretching or squeezing the points along the number line, preserving the shape of the z-score distribution according to the following formula:

$$X_{rescaled} = \frac{X - \min X}{\max X - \min X}$$

Normalized valence and arousal data complied with a normal distribution as confirmed by Shapiro–Wilk tests ($W = 0.97$ for valence data; $W = 0.97$ for arousal data).

Normalized data were then analyzed *via* two mixed-design ANOVA with avatar gender (male, female) and BP/dynamicity (low, middle, high) as within-subject factors and subject gender (male, female) as between-subject factor.

Results

Valence

Results of the rm ANOVA on valence ratings are illustrated in **Figure 4** (upper panels). A significant effect for the main factor BP was found [$F_{(2,104)} = 476.631$, $p < 0.001$, $\eta_p^2 = 0.902$]. Instead, neither the factor avatar [$F_{(1,52)} = 0.413$, $p = 0.523$, $\eta_p^2 = 0.008$] nor subject gender [$F_{(1,52)} = 0.014$, $p = 0.906$, $\eta_p^2 < 0.001$] showed a significant effect on

² Available on <https://renderpeople.com/free-3d-people/>.

³ <https://www.prolific.co/>

⁴ <https://pavlovio.org/>

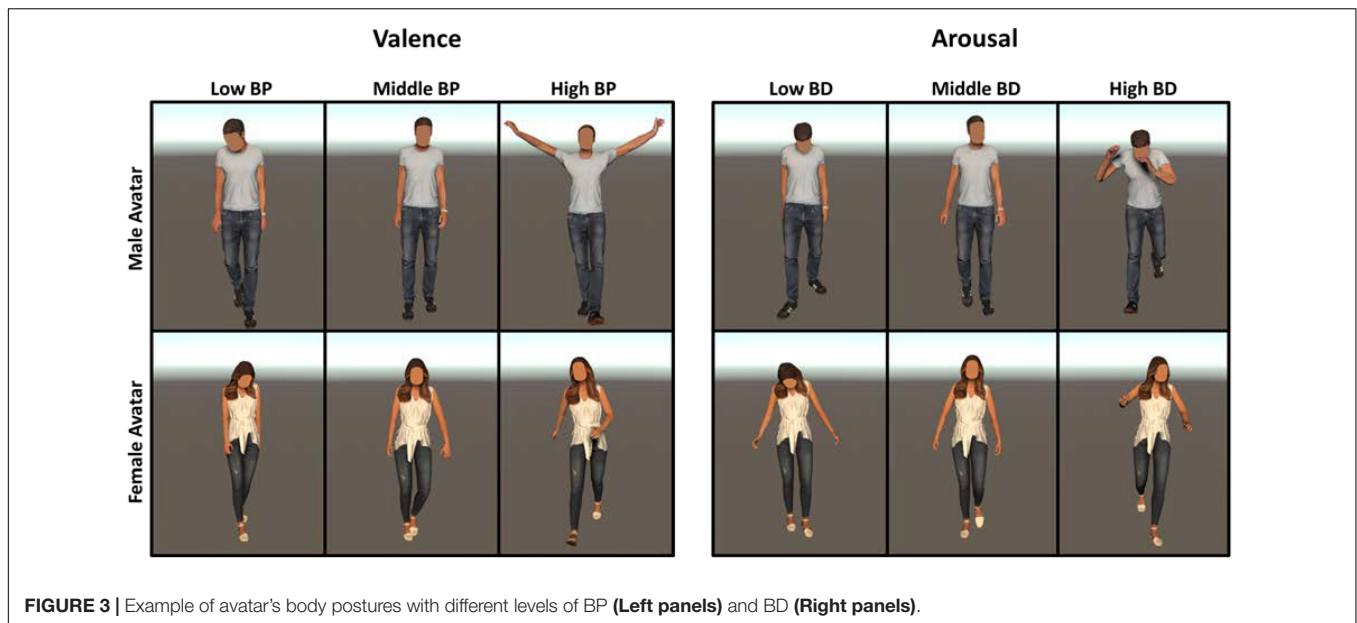


FIGURE 3 | Example of avatar's body postures with different levels of BP (Left panels) and BD (Right panels).

valence ratings. Bonferroni-corrected pairwise comparisons revealed that participants differentially judged avatars whose body postures belonged to different levels of BP (low < middle, $p < 0.001$; middle < high, $p < 0.001$; low < high, $p < 0.001$). Also, the interaction BP \times avatar was significant [$F_{(2,104)} = 35.615$, $p < 0.001$, $\eta_p^2 = 0.406$]. Specifically, Bonferroni-corrected pairwise comparison showed that female avatar with low level of BP was judged as less pleasant than male avatar with low level of BP ($p < 0.001$). Also, female avatar with high level of BP was perceived as more pleasant than male ones with high level of BP ($p < 0.001$).

Arousal

Figure 4 (lower panels) presents the results of the rm ANOVA on arousal ratings. We found a significant effect for the main factors BD [$F_{(2,104)} = 383.447$, $p < 0.001$, $\eta_p^2 = 0.881$] and avatar [$F_{(1,52)} = 15.073$, $p < 0.001$, $\eta_p^2 = 0.241$]. Conversely, the between-subject factor subject gender did not return a significant effect [$F_{(1,52)} = 0.367$, $p = 0.547$, $\eta_p^2 = 0.007$]. Bonferroni-corrected pairwise comparisons showed that differences among the levels of the main factor BD were all significant (low < middle, $p < 0.001$; middle < high, $p = 5.3 \times 10^{-7}$; low < high, $p < 0.001$). The significant main factor avatar showed that participants judged female avatar as more arousing than male ones. The interaction body dynamicity \times avatar [$F_{(2,104)} = 101.907$, $p < 0.001$, $\eta_p^2 = 0.662$] was also significant, revealing that female avatar with low level of BD was perceived as less arousing compared to male avatar with the same level of BD ($p < 0.001$). Also, female avatar was judged as more arousing than male ones when they belonged to the middle BD level ($p < 0.001$). Finally, we found no significant difference between arousal ratings provided on female avatar with a middle level of BD and female avatar with a high level of BD.

Experiment 1.B

Stimuli

A total of 180 videos were created reproducing the walking actions from which the postures of Experiment 1.A were originally extracted. Walking actions recorded on male (female) actors were used to animate a male (female) avatar, thus ensuring coherence between the gender of the actor and that of the virtual avatar.

Participants, Experimental Procedure, and Data Analysis

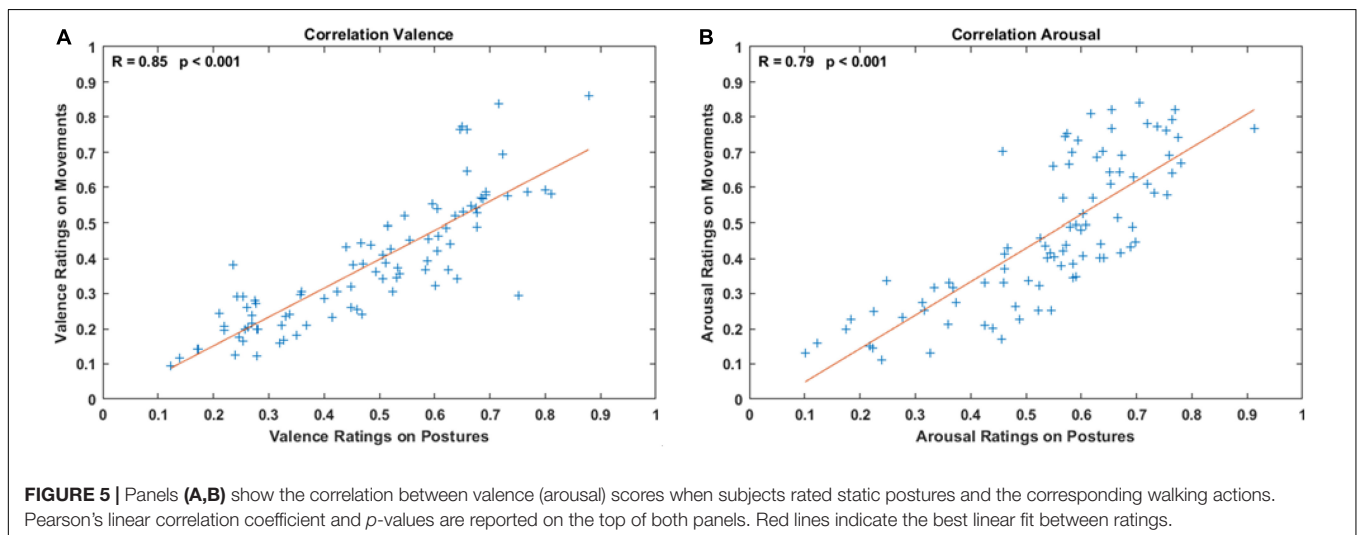
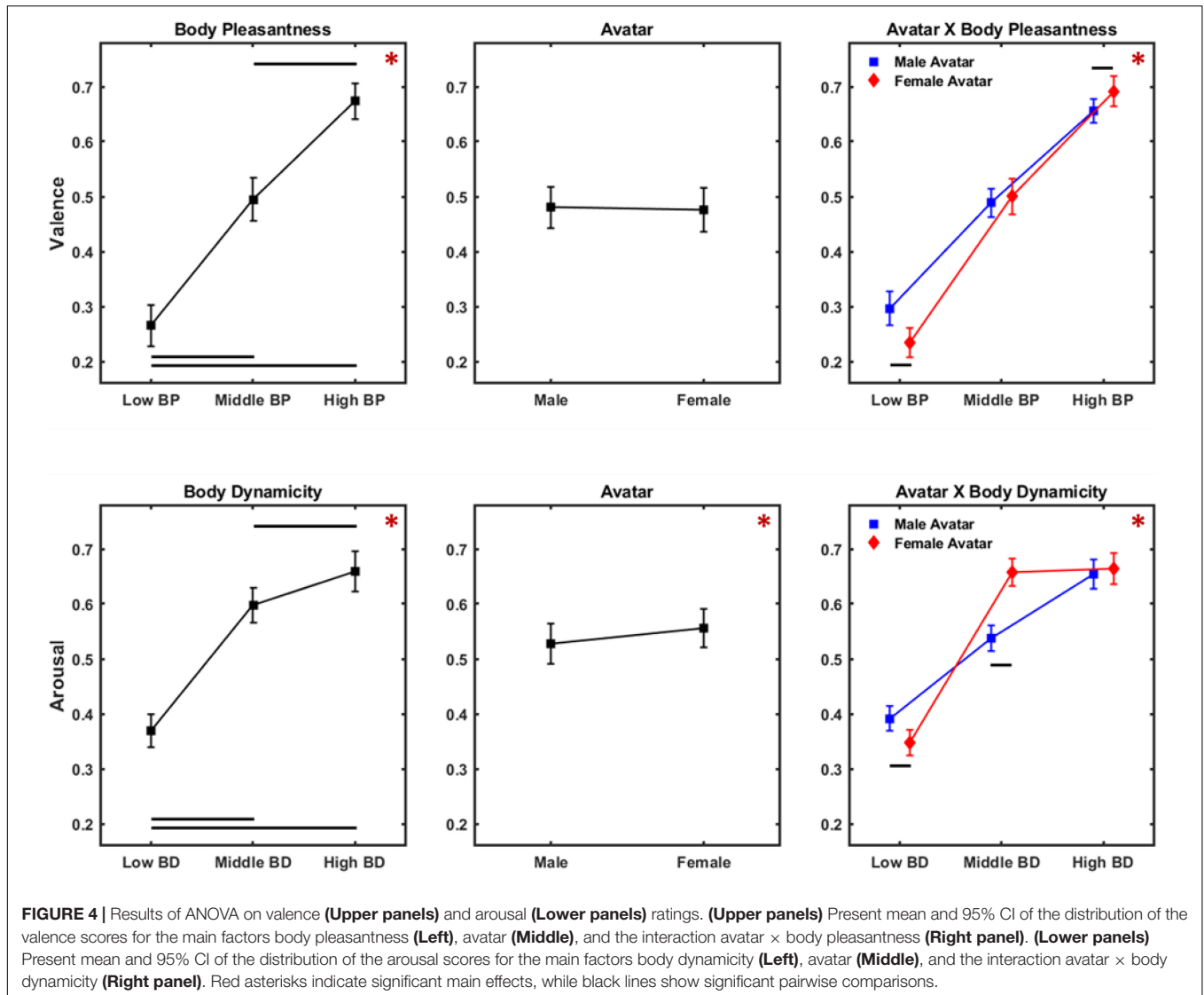
A total of 22 gender and age-matched participants were recruited and performed the experiment on Pavlovla [11 women aged 29.4 ± 4.3 years, 11 men aged 27.9 ± 4.6 years; two-sample t -test, $t_{(52)} = -0.766$, $p = 0.453$].

The adopted experimental procedure was the same as in Experiment 1.A, with the only difference that participants rated the valence and arousal perceived in the emotional walking instead of the corresponding representative static frame. Hence, on the left side of the screen, a video appeared for 2 s showing the walking action, while the two questions were presented on the right side with which participants could rate the arousal and valence level expressed by the avatar's walking through a VAS.

Valence and arousal ratings were normalized according to the same procedure used in the Experiment 1.A. Then, Pearson's linear correlation coefficient was computed to assess the correlation between valence (arousal) ratings that participants gave on static body postures and those given on the corresponding walking actions.

Results

Results of the correlation analysis are illustrated in **Figure 5**. Panel A shows the positive correlation between valence ratings between the two experiments ($R = 0.85$, $p < 0.001$; best linear



fit: $y = 0.82x - 0.01$). Similarly, panel B shows the correlation between arousal ratings ($R = 0.79$, $p < 0.001$; best linear fit: $y = 0.95x - 0.05$).

Discussion

We defined two body features related to pleasantness and dynamicity based on kinematic and postural information extracted from a set of emotional walking recorded with a motion capture system. We created a male and female avatar with body postures corresponding to three levels of BP and BD, coherently to the gender of the actor on which the kinematics were recorded. Experiment 1.A showed that participants coherently judged the avatar's bodily valence and arousal according to the defined BP and dynamicity levels. We also investigated whether the gender of the avatar influenced the perception of both valence and arousal, finding that participants perceived female avatar as more emotionally expressive than male ones. Experiment 1.B provided a ground truth comparison showing that valence and arousal levels perceived in the static body postures were consistent with those perceived in the corresponding walking actions.

Overall, we can argue that the defined bodily features allowed the affective information transfer from a full-body kinematic to a static body posture (see **Supplementary Material** for additional information). The adopted experimental procedure pointed to a difference in perception, possibly due to the avatar's gender. Despite the existing literature related to the gender difference in affective perception (Codispoti et al., 2008; Samadani et al., 2012), we did not observe any significant difference in the valence and arousal scores between male and female participants.

These results demonstrate that the defined BP, computed as the weighted sum of the openness of the body, leaning of the head, and leaning of the trunk, is a reliable descriptor of the valence dimension (Kleinsmith and Bianchi-Berthouze, 2007; Karg et al., 2013; Fourati and Pelachaud, 2018; Poyo Solanas et al., 2020), being able to transfer the affective information conveyed by an emotional walk to a static body posture. In fact, in this experiment, participants judged the valence level of the avatar coherently to the BP of the posture, distinguishing among low, middle, and high levels. In addition, the female avatar was rated in a more unpleasant state in the low BP condition and more pleasant in the high condition than the corresponding male avatar's judgments. Thus, participants judged the pleasantness state of the female avatar over a broader range, denoting that these postures were perceived as more expressive when characterized by unpleasant states and when expressing pleasant feelings. This result may reflect the female's higher emotional expressiveness. Because we normalized the displayed bodily features for the actor's gender, we may assume that the resulting differences depend on the observer's perception. Indeed, in line with biological and social models (Eagly and Wood, 1991; Brody and Hall, 2008), women are usually considered more emotionally expressive than men (Kring and Gordon, 1998; Kret and De Gelder, 2012; Chaplin and Aldao, 2013; Deng et al., 2016). When moving toward a virtual world, such biased perception is transferred from humans to virtual characters, thus making the female avatar seem to be more expressive when compared with

their male counterpart (DeWester et al., 2009; Zibrek et al., 2015; Bailey and Blackmore, 2017; Yang and Ryu, 2021).

As to the arousal dimension, we defined the BD as a parameter correlating with the velocity of the performed movement. Our results demonstrated that such affective information was transferred to static body postures. Indeed, participants coherently judged the three levels of BD, thus discriminating different arousal levels in the avatar's body posture. Previous research has shown that features such as velocity, acceleration, and jerk of the movement were highly correlated to the arousal content of emotional gaits, as well as of more specific movement such as drinking and knocking (Paterson et al., 2001; Pollick et al., 2001; Karg et al., 2010; McColl and Nejat, 2014). In automatic affect recognition, the quantity of motion is considered a discriminant factor to distinguish low-arousing movement from high-arousing ones (Castellano et al., 2007). Also, we found that arousal scores were significantly higher for female avatar than for males. Specifically, such biased perception depends on the higher arousal scores that participants gave female avatar with a middle level of BD. Indeed, these scores were comparable to those provided in the high condition. In addition, when compared with the male counterpart, female avatar was perceived as less arousing if characterized by low BD and more arousing if characterized by middle BD. We interpret such results as further evidence of the higher emotional expressivity of women (Grossman and Wood, 1993; Hess et al., 2000; Brody and Hall, 2008). However, this gender characteristic seems to be balanced in the high BD condition where the higher emotional expressivity of women is matched with the higher male tendency to show specific high-arousing emotions such as anger (Chaplin and Aldao, 2013). In fact, we found no differences in scores between male and female avatars in the high BD condition.

These findings reveal that the women's higher emotional expressiveness may involve both valence and arousal dimension, suggesting that participants perceive women as modulating the intensity of their affective states on a broader range than men.

To understand whether the women's higher emotional expressiveness could depend on the esthetics characteristics of the avatar used for the experiment or on the methodology we used to create the actual postures they assumed, we conducted a second experiment in which female avatar assumed body postures derived by male actors' kinematics and *vice versa*, thus creating an incoherent condition. Should participants still perceive female avatar as more emotionally expressive, we could argue that the avatar's esthetic characteristics (and not postures) mainly modulate the perceived emotional expressiveness. Conversely, should we find that female body postures are perceived more expressive even when assumed by a male avatar, we could conclude that female body postures (and not the avatar's esthetic) mainly contribute to modulate expressiveness levels.

EXPERIMENT 2

Stimuli

We used the same body postures of Experiment 1 to animate gender-opposite avatars to produce incoherence between the

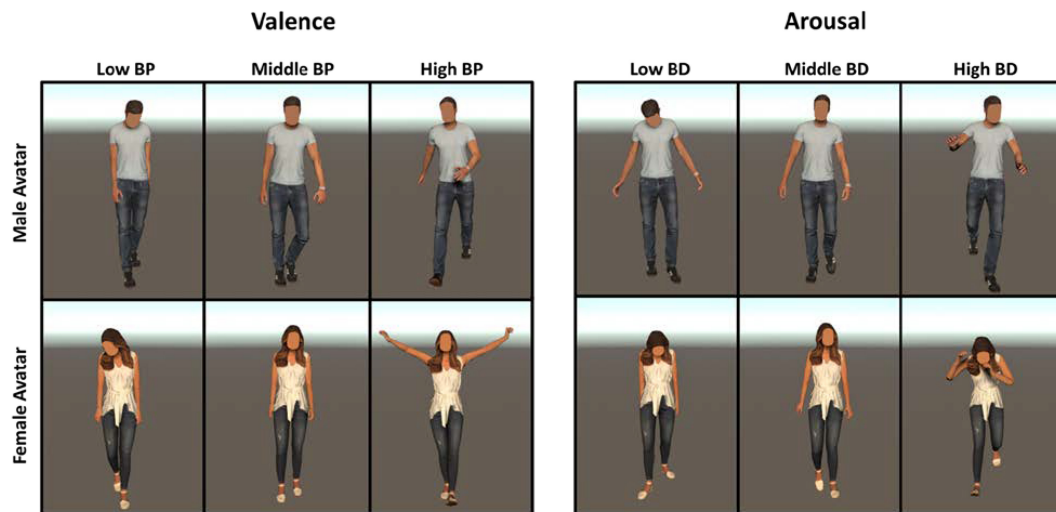


FIGURE 6 | Example of avatar's body postures with different levels of BP (**Right panels**) and BD (**Left panels**) used in Experiment 2. We assigned the body postures represented in **Figure 3** to the gender-opposite avatar.

gender of the avatar and that of the actor from which we extracted the body posture. Hence, body postures extracted from actions recorded on a male actor were assigned to a female avatar and *vice versa*. This procedure led to the creation of 180 stimuli, 90 characterized by different levels of BD and the other 90 by different levels of BP. In **Figure 6**, we illustrate incoherent stimuli with different levels of BP and BD, showing the same body postures of **Figure 3** represented by the gender-opposite avatar.

Participants, Experimental Procedure, and Data Analysis

The adopted experimental procedure was the same as in Experiment 1. A total of 54 age-matched participants [27 women aged 25.2 ± 5.6 years, 27 men aged 23.4 ± 3.6 years; two-sample *t*-test, $t_{(52)} = -1.202$, $p = 0.234$] were recruited through the online platform Prolific and then performed the experiment on Pavlovia. Also, participants were age-matched between Experiments 1 and 2 {two-way factorial ANOVA: no significant effect of the main factors subject gender [$F_{(1,104)} = 3.312$, $p = 0.072$] and experiment [$F_{(1,104)} = 0.793$, $p = 0.375$]}. As in Experiment 1, valence and arousal ratings were normalized between 0 and 1 and then analyzed *via* two mixed-design ANOVA with avatar (male, female) and BP/dynamicity (low, middle, high) as within-subject factors and subject gender (male female) as between-subject factor.

Results

Valence

Figure 7 (upper panels) shows results of rm ANOVA on valence scores. A significant effect for the main factors BP [$F_{(2,104)} = 570.417$, $p < 0.001$, $\eta_p^2 = 0.916$] and avatar [$F_{(1,52)} = 61.615$, $p = 2.2 \times 10^{-10}$, $\eta_p^2 = 0.542$] emerged. Conversely, no significant effect was observed for the main factor subject gender [$F_{(1,52)} = 1.857$, $p = 0.178$, $\eta_p^2 = 0.024$].

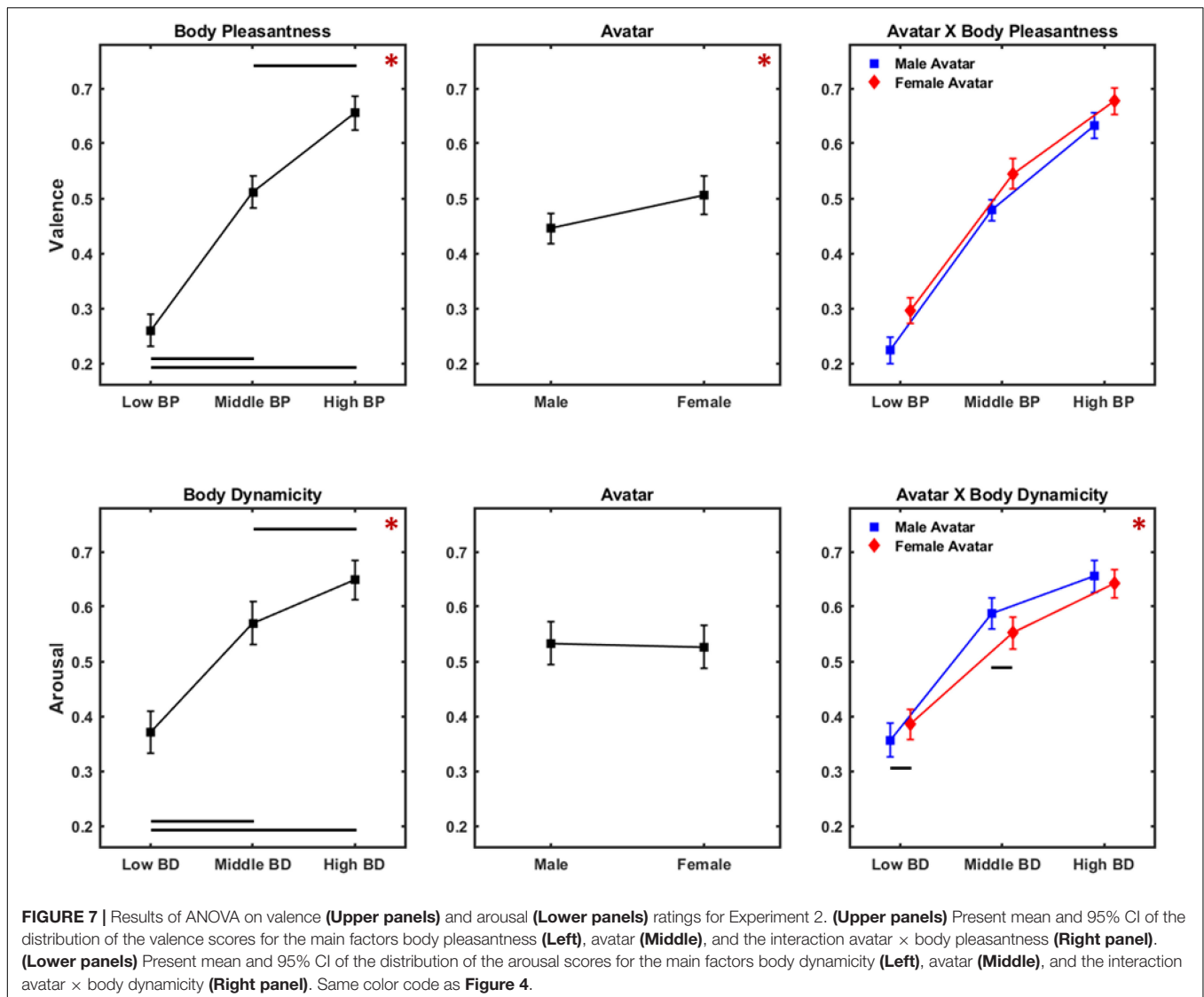
Bonferroni-corrected pairwise comparisons revealed that each level of BP received significantly different scores (low < middle, $p < 0.001$; middle < high, $p < 0.001$; low < high, $p < 0.001$).

Arousal

Figure 7 (lower panels) shows the results of rm ANOVA on arousal scores. A significant effect for the main factor BD [$F_{(2,104)} = 213.137$, $p < 0.001$, $\eta_p^2 = 0.804$] was observed, while both factors avatar [$F_{(1,52)} = 0.891$, $p = 0.349$, $\eta_p^2 = 0.017$] and subject gender [$F_{(1,52)} = 0.031$, $p = 0.860$, $\eta_p^2 < 0.001$] did not reveal a significant effect. Bonferroni-corrected pairwise comparisons showed that participants' ratings were significantly different for each level of the factor BD (low < middle, $p < 0.001$; middle < high, $p < 0.001$; low < high, $p < 0.001$). Also, the interaction avatar \times BD [$F_{(2,104)} = 13.304$, $p < 0.001$, $\eta_p^2 = 0.203$] was significant, revealing that female avatar with low BD was perceived as more arousing compared to male avatar with low BD ($p = 0.026$), and that male avatar with middle BD was perceived as more arousing compared to female avatar with middle BD ($p = 0.002$).

Discussion

In the second experiment, we disrupted the coherence between the actor's and avatar's gender to disentangle the relative contribution of body postures and avatars' esthetic characteristics in judgments on valence and arousal. Findings returned that participants distinguished three levels of valence and arousal expressed by the avatars' bodies according to the defined BP and dynamicity levels. Such results highlight that even if the avatar assumed postures recorded on actors of the opposite gender, the body features we defined could still transfer the affective information contained in the walking kinematic to a static body posture in terms of valence and arousal. Furthermore, we found that the combination of esthetic characteristics and body postures, i.e., the coherence between the actor's and avatar's



gender (and not singularly posture or esthetic factor), confers the higher emotional expressiveness to women.

As to the valence ratings, female avatar (with male postures) was perceived in a more pleasant state, regardless of the BP levels to which they belonged. However, considering the interaction between avatar gender and BP, the three levels of the body feature were similarly perceived between female and male avatars. Hence, disrupting the coherence between the gender of the avatar and that of the actor, we found that the avatar/actor gender incoherence spoils the difference between male and female avatars in terms of emotional expressiveness. A recent study reported a similar result using point-light stimuli representing emotional walking. The authors demonstrated that by depriving the participants of the structural cues of the walker gender stimuli were perceived as equally emotionally expressive (Halovic and Kroos, 2018b).

As to arousal, male avatar with body postures extracted from female actresses was perceived in a lower arousing state than

female avatar (with male postures) in the low BD condition. On the contrary, participants perceived the female avatar as less arousing than the male one in the middle condition. As in the first experiment, when considering the arousal dimension—regardless of the gender of the avatar—female postures were rated over a broader range compared to male ones. These results testify that the higher female expressiveness revealed in the arousal scores depends on gender-specific kinematic characteristics. These findings align with previous studies, where virtual puppets were more emotionally expressive when animated with female gestures (Yang and Ryu, 2021). However, these outcomes also show the importance of the avatar's esthetic characteristics, represented by the two genders, for arousal judgments. Participants judged the female body postures differently if applied to a male or female avatar, highlighting that also the gender of the avatar influences the subject's judgment on arousal perception. In fact, in the second experiment, participants distinguished three levels of BD for female avatar, while in the

first one, they confounded the middle with high BD. Similar results also emerged in previous research, showing that the same emotional postures were perceived as more emotionally expressive when represented by female virtual avatars (Zibrek et al., 2015; Cheng et al., 2020).

CONCLUSION

This study provides a new method that enables researchers to design body postures of virtual avatars with varying affective states, transferring affective information from dynamic walking to body postures. With this procedure, we created a set of virtual static stimuli potentially useful for studies exploring time-locked cognitive processes such as event-related potentials with electroencephalography, evoked potentials with transcranial magnetic stimulation, and reaction times in all those studies that aim to investigate the perception of emotional body postures.

Two online experiments proved the reliability of the proposed methodology and revealed that male and female avatars are differently perceived when their body posture derives from kinematics recorded on coherent or incoherent gender actors. Therefore, to prevent perceptual biases caused by individual characteristics, it is worth considering the “actor behind the avatar” when creating the virtual character with affective postures, i.e., the physical characteristics of the actor when transposing affective information to a virtual avatar. For instance, the height of the actor and his/her size could be relevant information that should also be considered to model the virtual avatar representative of the actor’s affective postures.

Further experiments could extend the validity of the presented methodology considering a more extensive set of kinematics comprising additional emotional gestures other than walking. Finally, exploiting 3D game engine software and virtual reality technologies, this methodology could be used in many experimental settings allowing researchers to resemble different social real-life situations to ultimately reach a deeper comprehension of how we perceive the affective states of others.

REFERENCES

- Abdi, H., and Williams, L. J. (2010). Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* 2, 433–459. doi: 10.1002/wics.101
- Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., and Young, A. W. (2004). Emotion Perception from Dynamic and Static Body Expressions in Point-Light and Full-Light Displays. *Perception* 33, 717–746. doi: 10.1068/p5096
- Aviezer, H., Trope, Y., and Todorov, A. (2012). Body Cues, Not Facial Expressions, Discriminate Between Intense Positive and Negative Emotions. *Science* 338, 1225–1229. doi: 10.1126/science.1224313
- Bailey, J. D., and Blackmore, K. L. (2017). “Gender and the perception of emotions in avatars,” in *Proceedings of the Australasian Computer Science Week Multiconference*, (Geelong Australia: ACM), 1–8. doi: 10.1145/3014812.3014876
- Barliya, A., Omlor, L., Giese, M. A., Berthoz, A., and Flash, T. (2013). Expression of emotion in the kinematics of locomotion. *Exp. Brain Res.* 225, 159–176. doi: 10.1007/s00221-012-3357-4
- Bernardet, U., Alaoui, S. F., Studd, K., Bradley, K., Pasquier, P., and Schiphorst, T. (2019). Assessing the reliability of the Laban Movement Analysis system. *PLoS One* 14:e0218179. doi: 10.1371/journal.pone.0218179

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Comitato Etico AVEN. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

PP, DR, and GV contributed to the conception and design of the study. PP and GMG collected the data. PP performed the statistical analysis and wrote the first draft of the manuscript. GMG, PA, and FC wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

GV has received research support from Lombardini22. This study was supported by a research agreement between Lombardini22 and IN-CNR.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2022.842433/full#supplementary-material>

- Bernhardt, D., and Robinson, P. (2007). “Detecting Affect from Non-stylised Body Motions,” in *Affective Computing and Intelligent Interaction Lecture Notes in Computer Science*, eds A. C. R. Paiva, R. Prada, and R. W. Picard (Berlin: Springer), 59–70. doi: 10.1007/978-3-540-74889-2_6
- Bhattacharya, U., Rewkowski, N., Guhan, P., Williams, N. L., Mittal, T., Bera, A., et al. (2020). “Generating Emotive Gaits for Virtual Agents Using Affect-Based Autoregression,” in *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* Brazil: IEEE, 24–35. doi: 10.1109/ISMAR50242.2020.00020
- Brody, L., and Hall, J. (2008). “Gender and emotion in context,” in *Handbook of Emotions: third edition*, eds M. Lewis, J. M. Haviland-Jones, and L. F. Barrett (New York, NY: The Guilford Press), 395–408.
- Buisine, S., Courgeon, M., Charles, A., Clavel, C., Martin, J.-C., Tan, N., et al. (2014). The Role of Body Postures in the Recognition of Emotions in Contextually Rich Scenarios. *Int. J. Hum. Comput. Interact.* 30, 52–62. doi: 10.1080/10447318.2013.802200
- Castellano, G., Villalba, S. D., and Camurri, A. (2007). “Recognising Human Emotions from Body Movement and Gesture Dynamics,” in *Affective Computing and Intelligent Interaction Lecture Notes in Computer Science*, eds

- A. C. R. Paiva, R. Prada, and R. W. Picard (Berlin: Springer), 71–82. doi: 10.1007/978-3-540-74889-2_7
- Chaplin, T. M., and Aldao, A. (2013). Gender Differences in Emotion Expression in Children: A Meta-Analytic Review. *Psychol. Bull.* 139, 735–765. doi: 10.1037/a0030737
- Cheng, J., Zhou, W., Lei, X., Adamo, N., and Benes, B. (2020). “The Effects of Body Gestures and Gender on Viewer’s Perception of Animated Pedagogical Agent’s Emotions,” in *Human-Computer Interaction. Multimodal and Natural Interaction Lecture Notes in Computer Science*, ed. M. Kurosu (Cham: Springer International Publishing), 169–186. doi: 10.1007/978-3-030-49062-1_11
- Clavel, C., Plessier, J., Martin, J.-C., Ach, L., and Morel, B. (2009). “Combining Facial and Postural Expressions of Emotions in a Virtual Character,” in *Intelligent Virtual Agents. IVA 2009. Lecture Notes in Computer Science*, 5773, eds Z. Ruttkay, M. Kipp, A. Nijholt, and H. H. Vilhjálmsson (Berlin: Springer), doi: 10.1007/978-3-642-04380-2_31
- Codispoti, M., Surcinelli, P., and Baldaro, B. (2008). Watching emotional movies: Affective reactions and gender differences. *Int. J. Psychophysiol.* 69, 90–95. doi: 10.1016/j.ijpsycho.2008.03.004
- Cohn, J. F., Ambadar, Z., and Ekman, P. (2007). “Observer-based measurement of facial expression with the Facial Action Coding System,” in *handbook of Emotion Elicitation and Assessment Series in Affective Science*, eds J. A. Coan and J. J. B. Allen (New York, NY: Oxford University Press), 203–221.
- Colombetti, G. (2005). Appraising Valence. *J. Conscious. Stud.* 12, 103–126.
- Coulson, M. (2004). Attributing Emotion to Static Body Postures: Recognition Accuracy, Confusions, and Viewpoint Dependence. *J. Nonverbal Behav.* 28, 117–139. doi: 10.1023/B:JONB.0000023655.25550.be
- Crenn, A., Khan, R. A., Meyer, A., and Bouakaz, S. (2016). “Body expression recognition from animated 3D skeleton,” in *2016 International Conference on 3D Imaging (IC3D)*, (Belgium: Université de Lyon), 1–7. doi: 10.1109/IC3D.2016.7823448
- Dael, N., Goudbeek, M., and Scherer, K. R. (2013). Perceived Gesture Dynamics in Nonverbal Expression of Emotion. *Perception* 42, 642–657. doi: 10.1068/p7364
- Dael, N., Mortillaro, M., and Scherer, K. R. (2012). The Body Action and Posture Coding System (BAP): Development and Reliability. *J. Nonverbal Behav.* 36, 97–121. doi: 10.1007/s10919-012-0130-0
- de Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nat. Rev. Neurosci.* 7, 242–249. doi: 10.1038/nrn1872
- de Gelder, B. (2009). Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 364, 3475–3484. doi: 10.1098/rstb.2009.0190
- de Gelder, B., Van den Stock, J., Meeren, H. K. M., Sinke, C. B. A., Kret, M. E., and Tamietto, M. (2010). Standing up for the body. Recent progress in uncovering the networks involved in the perception of bodies and bodily expressions. *Neurosci. Biobehav. Rev.* 34, 513–527. doi: 10.1016/j.neubiorev.2009.10.008
- De Silva, P. R., and Bianchi-Berthouze, N. (2004). Modeling human affective postures: an information theoretic characterization of posture features. *Comput. Animat. Virtual Worlds* 15, 269–276. doi: 10.1002/cav.29
- Deligianni, F., Guo, Y., and Yang, G.-Z. (2019). From Emotions to Mood Disorders: A Survey on Gait Analysis Methodology. *IEEE J. Biomed. Health Inform.* 23, 2302–2316. doi: 10.1109/JBHI.2019.2938111
- Deng, Y., Chang, L., Yang, M., Huo, M., and Zhou, R. (2016). Gender Differences in Emotional Response: Inconsistency between Experience and Expressivity. *PLoS One* 11:e0158666. doi: 10.1371/journal.pone.0158666
- DeWester, D., Nah, F. F.-H., Gervais, S. J., and Siau, K. (2009). “Are Male and Female Avatars Perceived Equally in 3d Virtual Worlds?,” in *AMCIS 2009 Proceedings*, (California: Association for Information Systems), 386.
- Eagly, A. H., and Wood, W. (1991). Explaining Sex Differences in Social Behavior: A Meta-Analytic Perspective. *Pers. Soc. Psychol. Bull.* 17, 306–315. doi: 10.1177/0146167291173011
- Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146
- Fourati, N., and Pelachaud, C. (2014). “Emily: Emotional body expression in daily actions database,” in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14)*, (Reykjavik: European Language Resources Association (ELRA)), 3486–3493.
- Fourati, N., and Pelachaud, C. (2015). “Relevant body cues for the classification of emotional body expression in daily actions,” in *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, (China: IEEE), 267–273. doi: 10.1109/ACII.2015.7344582
- Fourati, N., and Pelachaud, C. (2018). Perception of Emotions and Body Movement in the Emily Database. *IEEE Trans. Affect. Comput.* 9, 90–101. doi: 10.1109/TAFFC.2016.2591039
- Gross, M. M., Crane, E. A., and Fredrickson, B. L. (2012). Effort-Shape and kinematic assessment of bodily expression of emotion during gait. *Hum. Mov. Sci.* 31, 202–221. doi: 10.1016/j.humov.2011.05.001
- Grossman, M., and Wood, W. (1993). Sex differences in intensity of emotional experience: A social role interpretation. *J. Pers. Soc. Psychol.* 65, 1010–1022. doi: 10.1037/0022-3514.65.5.1010
- Halovic, S., and Kroos, C. (2018a). Not all is noticed: Kinematic cues of emotion-specific gait. *Hum. Mov. Sci.* 57, 478–488. doi: 10.1016/j.humov.2017.11.008
- Halovic, S., and Kroos, C. (2018b). Walking my way? Walker gender and display format Confounds the perception of specific emotions. *Hum. Mov. Sci.* 57, 461–477. doi: 10.1016/j.humov.2017.10.012
- Hess, U., Senécal, S., Kirouac, G., Herrera, P., Philippot, P., and Kleck, R. E. (2000). Emotional expressivity in men and women: Stereotypes and self-perceptions. *Cogn. Emot.* 14, 609–642. doi: 10.1080/02699930050117648
- Hicheur, H., Kadone, H., Grèzes, J., and Berthoz, A. (2013). “Perception of Emotional Gaits Using Avatar Animation of Real and Artificially Synthesized Gaits,” in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, (Switzerland: IEEE)460–466. doi: 10.1109/ACII.2013.82
- Karg, M., Kühnlenz, K., and Buss, M. (2010). Recognition of Affect Based on Gait Patterns. *IEEE Trans. Syst. Man Cybern. Syst.* 40, 1050–1061. doi: 10.1109/TSMCB.2010.2044040
- Karg, M., Samadani, A.-A., Gorbet, R., Kühnlenz, K., Hoey, J., and Kulić, D. (2013). Body Movements for Affective Expression: A Survey of Automatic Recognition and Generation. *IEEE Trans. Affect. Comput.* 4, 341–359. doi: 10.1109/T-AFFC.2013.29
- Kleinsmith, A., and Bianchi-Berthouze, N. (2007). “Recognizing Affective Dimensions from Body Posture,” in *Affective Computing and Intelligent Interaction Lecture Notes in Computer Science*, eds A. C. R. Paiva, R. Prada, and R. W. Picard (Berlin: Springer), 48–58. doi: 10.1007/978-3-540-74889-2_5
- Kleinsmith, A., and Bianchi-Berthouze, N. (2013). Affective Body Expression Perception and Recognition: A Survey. *IEEE Trans. Affect. Comput.* 4, 15–33. doi: 10.1109/T-AFFC.2012.16
- Kleinsmith, A., De Silva, P. R., and Bianchi-Berthouze, N. (2006). Cross-cultural differences in recognizing affect from body posture. *Interact. Comput.* 18, 1371–1389. doi: 10.1016/j.intcom.2006.04.003
- Kragel, P. A., and LaBar, K. S. (2016). Decoding the Nature of Emotion in the Brain. *Trends Cogn. Sci.* 20, 444–455. doi: 10.1016/j.tics.2016.03.011
- Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biol. Psychol.* 84, 394–421. doi: 10.1016/j.biopsycho.2010.03.010
- Kret, M. E., and de Gelder, B. (2010). Social context influences recognition of bodily expressions. *Exp. Brain Res.* 203, 169–180. doi: 10.1007/s00221-010-2220-8
- Kret, M. E., and De Gelder, B. (2012). A review on sex differences in processing emotional signals. *Neuropsychologia* 50, 1211–1221. doi: 10.1016/j.neuropsychologia.2011.12.022
- Kring, A. M., and Gordon, A. H. (1998). Sex differences in emotion: Expression, experience, and physiology. - *PsychNET. J. Pers. Soc. Psychol.* 74, 686–703. doi: 10.1037//0022-3514.74.3.686
- Kuppens, P., Tuerlinckx, F., Russell, J. A., and Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychol. Bull.* 139, 917–940. doi: 10.1037/a0030811
- Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., and Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behav. Brain Sci.* 35, 121–143. doi: 10.1017/S0140525X11000446
- McColl, D., and Nejat, G. (2014). “Determining the affective body language of older adults during socially assistive HRI,” in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, (Chicago: IEEE), 2633–2638. doi: 10.1109/IROS.2014.6942922
- McHugh, J. E., McDonnell, R., O’Sullivan, C., and Newell, F. N. (2010). Perceiving emotion in crowds: the role of dynamic body postures on the perception of emotion in crowded scenes. *Exp. Brain Res.* 204, 361–372. doi: 10.1007/s00221-009-2037-5

- Montepare, J. M., Goldstein, S. B., and Clausen, A. (1987). The identification of emotions from gait information. *J. Nonverbal. Behav.* 11, 33–42. doi: 10.1007/BF00999605
- Nakagawa, K., Shinozawa, K., Ishiguro, H., Akimoto, T., and Hagita, N. (2009). “Motion modification method to control affective nuances for robots,” in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, (Louis: IEEE), 5003–5008. doi: 10.1109/IROS.2009.5354205
- Palan, S., and Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *J. Behav. Exp. Finance* 17, 22–27. doi: 10.1016/j.jbef.2017.12.004
- Paterson, H. M., Pollick, F. E., and Sanford, A. J. (2001). “The Role of Velocity in Affect Discrimination,” in *Proceedings of the Annual Meeting of the Cognitive Science Society* 23. Available Online at: <https://escholarship.org/uc/item/3191m9bh> (accessed on Dec 6, 2021).
- Pollick, F. E., Paterson, H. M., Bruderlin, A., and Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition* 82, B51–B61. doi: 10.1016/S0010-0277(01)00147-0
- Poyo Solanas, M., Vaessen, M., and Gelder, B. (2020). Computation-based feature representation of body expressions in the human brain. *Cereb. Cortex* 30, 6376–6390. doi: 10.1093/cercor/bhaa196
- Presti, P., Ruzzon, D., Avanzini, P., Caruana, F., Rizzolatti, G., and Vecchiato, G. (2021). Dynamic experience of architectural forms affects arousal and valence perception in virtual environments. *Res. Square* [Preprint]. doi: 10.21203/rs.3.rs-910384/v1
- Randhavane, T., Bera, A., Kapsaskis, K., Gray, K., and Manocha, D. (2019a). FVA: Modeling Perceived Friendliness of Virtual Agents Using Movement Characteristics. *IEEE Trans. Vis. Comput. Graph.* 25, 3135–3145. doi: 10.1109/TVCG.2019.2932235
- Randhavane, T., Bera, A., Kapsaskis, K., Sheth, R., Gray, K., and Manocha, D. (2019b). “EVA: Generating Emotional Behavior of Virtual Agents using Expressive Features of Gait and Gaze,” in *ACM Symposium on Applied Perception 2019 SAP '19*, (New York, NY: Association for Computing Machinery), 1–10. doi: 10.1145/3343036.3343129
- Randhavane, T., Bera, A., Kubin, E., Gray, K., and Manocha, D. (2021). Modeling Data-Driven Dominance Traits for Virtual Characters Using Gait Analysis. *IEEE Trans. Vis. Comput. Graph.* 27, 2967–2979. doi: 10.1109/TVCG.2019.2953063
- Roether, C. L., Omlor, L., Christensen, A., and Giese, M. A. (2009). Critical features for the perception of emotion from gait. *J. Vis.* 9:15. doi: 10.1167/9.6.15
- Samadani, A.-A., Gorbet, R., and Kulić, D. (2012). “Gender Differences in the Perception of Affective Movements,” in *Human Behavior Understanding Lecture Notes in Computer Science*, eds A. A. Salah, J. Ruiz-del-Solar, Ç Meriçli, and P.-Y. Oudeyer (Berlin: Springer), 65–76. doi: 10.1007/978-3-642-34014-7_6
- Sanchez-Vives, M. V., and Slater, M. (2005). From presence to consciousness through virtual reality. *Nat. Rev. Neurosci.* 6, 332–339. doi: 10.1038/nrn1651
- Sanghvi, J., Castellano, G., Leite, I., Pereira, A., McOwan, P. W., and Paiva, A. (2011). “Automatic analysis of affective postures and body motion to detect engagement with a game companion,” in *Proceedings of the 6th international conference on Human-robot interaction HRI '11*, (New York, NY: Association for Computing Machinery), 305–312. doi: 10.1145/1957656.1957781
- Slater, M., and Sanchez-Vives, M. V. (2016). Enhancing Our Lives with Immersive Virtual Reality. *Front. Robot. AI* 3:74. doi: 10.3389/frobt.2016.00074
- Stephens-Fripp, B., Naghdy, F., Stirling, D., and Naghdy, G. (2017). Automatic Affect Perception Based on Body Gait and Posture: A Survey. *Int. J. Soc. Robot.* 9, 617–641. doi: 10.1007/s12369-017-0427-6
- Venture, G., Kadone, H., Zhang, T., Grèzes, J., Berthoz, A., and Hicheur, H. (2014). Recognizing Emotions Conveyed by Human Gait. *Int. J. Soc. Robot.* 6, 621–632. doi: 10.1007/s12369-014-0243-1
- Vinayagamoorthy, V., Steed, A., and Slater, M. (2008). The Impact of a Character Posture Model on the Communication of Affect in an Immersive Virtual Environment. *IEEE Trans. Vis. Comput. Graph.* 14, 965–982. doi: 10.1109/TVCG.2008.62
- Yang, E., and Ryu, J. (2021). Do We Perceive Emotional Gender Gesture of Virtual Avatar As Intended? *Int. J. Educ. Technol.* 15, 129–139.
- Zhao, N., Zhang, Z., Wang, Y., Wang, J., Li, B., Zhu, T., et al. (2019). See your mental state from your walk: Recognizing anxiety and depression through Kinect-recorded gait data. *PLoS One* 14:e0216591. doi: 10.1371/journal.pone.0216591
- Zibrek, K., Hoyet, L., Ruhland, K., and McDonnell, R. (2015). Exploring the Effect of Motion Type and Emotions on the Perception of Gender in Virtual Humans. *ACM Trans. Appl. Percept.* 11:20. *ACM Trans. Appl. Percept.* doi: 10.1145/2767130

Conflict of Interest: DR was employed by company TUNED, Lombardini22.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Presti, Ruzzon, Galasso, Avanzini, Caruana and Vecchiato. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



OPEN ACCESS

EDITED BY

Anil Ufuk Batmaz,
Kadir Has University, Turkey

REVIEWED BY

Jiawei Zhou,
Wenzhou Medical University, China
JunYun Zhang,
Peking University, China

*CORRESPONDENCE

Longqian Liu
b.q15651@hotmail.com

†These authors have contributed
equally to this work and share the first
authorship

SPECIALTY SECTION

This article was submitted to
Perception Science,
a section of the journal
Frontiers in Neuroscience

RECEIVED 23 January 2022

ACCEPTED 29 June 2022

PUBLISHED 15 July 2022

CITATION

Zhu B, Liao M and Liu L (2022)
Measuring the impact of suppression
on visual acuity in children with
amblyopia using a dichoptic visual
acuity chart.
Front. Neurosci. 16:860620.
doi: 10.3389/fnins.2022.860620

COPYRIGHT

© 2022 Zhu, Liao and Liu. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Measuring the impact of suppression on visual acuity in children with amblyopia using a dichoptic visual acuity chart

Bixia Zhu^{1†}, Meng Liao^{2†} and Longqian Liu^{1,2*}

¹Department of Optometry and Visual Science, West China Hospital, Sichuan University, Chengdu, China, ²Department of Ophthalmology, West China Hospital, Sichuan University, Chengdu, China

Purpose: To develop a novel dichoptic visual acuity chart that measures the impact of interocular suppression on the visual acuity of each eye when two eyes are open.

Methods: Fifty-four subjects (19 anisometropic amblyopia, 20 treated amblyopia, and 15 normal children) participated in this study. The visual acuity that was tested under dichoptic-optotypes condition (i.e., presented optotypes to the untested eye) was compared with that under monocular condition (i.e., cover the untested eye with opaque patch). Visual acuity differences between these two conditions were compared among the three groups. The correlations between visual acuity differences and the depth of interocular suppression were then computed. Some participants performed the visual acuity test under dichoptic-luminance condition (i.e., presented mean luminance to the untested eye), and the test-retest reliability was established.

Results: A reduced visual acuity of the non-dominant eye was found in the dichoptic-optotypes condition for the amblyopia group ($P < 0.001$) and the treated group ($P = 0.001$); the difference in the treated group was less than that in the amblyopia group ($P < 0.001$) but more than that in the normal group ($P = 0.026$). A significant correlation was found between the visual acuity differences and the depth of suppression, which was tested with a binocular phase combination task ($P = 0.005$). No change was found in the dichoptic-luminance condition.

Conclusion: The amblyopic eye and the previous amblyopic eye seem to suffer from a reduced visual acuity when two eyes are open due to suppression. This was successfully captured by our novel and reliable dichoptic-optotypes visual acuity chart.

KEYWORDS

amblyopia, suppression, visual acuity, dichoptic eye chart, mean luminance

Introduction

Amblyopia is a neurodevelopmental disorder that results from poor visual development during the critical period. Symptoms include poor monocular visual acuity and impaired binocular function. The rate of amblyopia in the general population is 1–4% (Multi-ethnic Pediatric Eye Disease Study, 2008; Williams et al., 2008; Multi-Ethnic Pediatric Eye Disease Study, 2009); the number of individuals with amblyopia might be 221.9 million by 2040 (Fu et al., 2019). In unilateral amblyopes, there is an imbalanced suppression of visual input between the eyes; for example, the suppression from the fellow eye to the amblyopic eye is stronger than the one originating from the amblyopic eye to the fellow eye, thereby creating an imbalance (Huang et al., 2011; Zhou et al., 2018). Studies indicate this form of imbalanced suppression between the eyes in amblyopia determines both monocular and binocular visual functions (Hess and Thompson, 2015). However, the standard means to diagnose amblyopia is measuring the lines of logMAR difference in visual acuity between the eyes (Wallace et al., 2018). Also, monocular visual acuity of the tested eye is usually tested while the untested eye is occluded. However, the monocular occlusion minimizes interocular interaction (Lai et al., 2012; Jia et al., 2015).

It is believed that imbalanced suppression between the eyes perturbs the visual acuity of the amblyopic eye. This has been shown in previous studies. For instance, Pugh (1954) used the orthoptoscope to present test dots (subtending different angles equivalent to 1/60–6/6 Snellen letter) to the amblyopic eye and fixation dot to the fellow eye. By changing the luminance of the fellow eye via neutral density filters, Pugh showed that the acuity of the amblyopic eye decreased as the light level of the fellow eye increased. Moreover, von Noorden and Leffler (1966) also found that the visual acuity of the strabismic amblyopic eye was worse when there was visual input in the fellow eye. von Noorden and Leffler (1966) used polaroid filters to present the visual acuity chart to the amblyopic eye, but a black chart surface to the fellow eye. Lai et al. (2011) and Lai et al. (2012) showed that the visual acuity of the amblyopic eye was reduced when it was partially patched (by a square patch that occluded the central visual field of the fellow eye) compared to when it was fully patched. Nevertheless, the relationship between the content of reduced visual acuity in the amblyopic eye under the dichoptic condition and the depth of suppression seems to remain opaque.

Recovery of amblyopia is often determined by tracking the difference in visual acuity between the eyes after a period of monocular treatment. However, studies show that individuals who have been supposedly treated with amblyopia as measured with their improved visual acuity of the amblyopic eye still exhibit binocular imbalance as a function of spatial frequency (Chen et al., 2017, 2021; Zhao et al., 2017). This finding indicates that the binocular imbalance, which could be due to imbalanced suppression, in amblyopia still remains even if visual acuity gets

improved throughout standard treatment such as monocular occlusion of the fellow eye (Kehrein et al., 2016; Jia et al., 2018; Chen et al., 2019). If suppression plays a primary role and an impaired visual acuity is merely a subsequent event due to suppression (Li et al., 2011; Hess et al., 2014), residual binocular imbalance indicates that the current method of treatment for amblyopia is inadequate to ensure a full recovery of the visual function. Also, to what content the binocular imbalance perturbs the visual acuity of the previous amblyopic eye while the fellow eye receives visual input is still unclear.

To answer our question, we designed a dichoptic visual acuity chart. This chart has two new features. First, suppression has been found to exhibit dependence on spatial frequency (Kwon et al., 2015; Mao et al., 2020), and that it can be influenced by interocular contrast (Birch et al., 2019) or luminance (Zhou et al., 2013) difference. In other words, the presentation of stimuli at different spatial frequencies, contrast, or luminance between both eyes might introduce interocular imbalance that might otherwise be absent. However, in our study, we presented the untested eye at the same spatial frequencies (i.e., the same size of optotypes), contrast (i.e., 100% Weber Contrast) and luminance as the tested eye. Second, the optotypes shown to the two eyes were vertically arranged and were not perceived as being overlapped; this feature is in contrast as those used in previous studies where overlapping optotypes were used to test suppression at various interocular contrast ratios (Kwon et al., 2015; Birch et al., 2016). In these studies, optotypes presented to two eyes should be of low spatial frequencies so that subjects can see them clearly. When it comes to a visual acuity test (higher spatial frequency), overlapping arrangement could produce confusion to the observer.

We found that the amblyopic eye and the previous amblyopic eye had reduced visual acuity when the fellow eye was viewing optotypes rather than mean luminance. The magnitude of visual acuity change in dichoptic and monocular conditions was correlated with the depth of suppression. Our new dichoptic letter chart demonstrated a robust test-retest reliability. Therefore, we recommend that the dichoptic visual acuity chart be used to measure the visual acuity of amblyopes in the future.

Materials and methods

Participants

Fifty-four children were enrolled in the ophthalmology department of the Western China Hospital, Sichuan University: 15 normal individuals (9.27 ± 2.19 years old; mean \pm SD), 19 anisometric amblyopes (8.95 ± 2.97 years old), and 20 treated amblyopes (8.00 ± 2.73 years old). All participants underwent comprehensive clinical examinations, including previous treatment history, best-corrected visual acuity (BCVA),

slit-lamp examination, ophthalmoscopic exam, stereoacuity, alignment exam, and extraocular muscle movements. BCVA was tested using a Tumbling E Logarithmic Visual Acuity Chart (xk100-06, China). Stereoacuity was tested with the TNO stereogram (TNO 18th, Lameris Ootech BV, Celsiusbaan 6B, 3439 NC, Nieuwegein, the Netherlands). This study adhered to the tenets of the Declaration of Helsinki and was approved by the ethics committee of the Western China Hospital. Written informed consent was obtained from the patients' guardians or parents.

Amblyopia was diagnosed according to the Preferred Practice Pattern of The American Academy of Ophthalmology (Wallace et al., 2018). Individuals were classified as having anisometropic amblyopia if they had an interocular BCVA difference greater than 2 lines, or interocular BCVA difference less than 2 lines but the amblyopic eye's visual acuity worse than 0.1 logMAR, with anisometropia greater than 1.50 D in spherical lens or 1.00 D in cylinder lens. Treated amblyopia was defined as a BCVA of the previous amblyopic eye achieving 0.1 logMAR and an interocular acuity difference of less than 2 lines. The normal controls had a normal BCVA (≤ 0.1 logMAR), no risk factors (i.e., strabismus, uncorrected anisometropia), and no history of amblyopia. Patients were excluded from this study if they had a history of organic eye disease and had undergone patching or cycloplegia within 4 h just before the measurement of our experiments. The clinical details of the participants are provided in [Supplementary Table 1](#).

In this report, we refer to the amblyopic eye of the amblyopia group, the previous amblyopic eye of the treated group, and the non-acuity dominant eye (i.e., the eye with worse BCVA) (Coren and Kaplan, 1973; Vedamurthy et al., 2007) of the normal group as the non-dominant eye (NDE) and the other eye as the dominant eye (DE).

Apparatus

The dichoptic visual acuity test and the binocular phase combination task were conducted using MATLAB 2017b (The Mathworks, Inc., Natick, MA, United States) with PsychToolBox 3.0.14 on a gamma-corrected polarized 3D monitor (27-in; D2757PH, AOC, Inc., $1,920' \times 1,080'$) in a dark room. The refresh rate was 60 Hz. Polarized glasses were used during the test. The maximum luminance was set to 96.4 cd/m^2 and was reduced to 44.8 cd/m^2 using polarized glasses.

Experimental design

We performed two experiments in our study. In Experiment 1, we tested the visual acuity of each eye in dichoptic and monocular conditions using the dichoptic letter chart to measure their changes in visual acuity (CVA). Then, we

examined the relationship between visual acuity changes in the depth of suppression by using a binocular phase combination task, which measures the relative contribution of each eye in binocular vision. However, while the tested eye viewed the optotypes in the dichoptic-optotypes condition of this study, the corresponding region of the untested eye was presented with mean luminance. The impact of presenting mean luminance to the fellow eye on the visibility of the amblyopic eye has been under dispute (Huang et al., 2012; Zhou et al., 2014; Jia et al., 2015). Thus, in Experiment 2, we examined the effect of mean luminance on the visual acuity change by testing the visual acuity of each eye in dichoptic-optotypes, dichoptic-luminance, and monocular conditions. To achieve a better understanding of the relationship between suppression and CVA, we measured suppression using a Worth 4-dot test. Finally, we evaluated the test-retest reliability of the dichoptic letter chart.

Dichoptic visual acuity test

As shown in [Figure 1](#), the dichoptic letter chart was comprised of a striped envelope, four short lines and a line of E letters; these were presented to the tested eye on a background of 96.4 cd/m^2 . The contrast was fixed at 100% throughout the test. The size of the letter E was drawn in logMAR form and was adjusted in a step of 0.1 logMAR through a keyboard. The test range was 1.0 logMAR to -0.1 logMAR. In Experiment 1, each row had four letters with one letter referring to 0.025 logMAR (see [Figure 1A](#)). In Experiment 2, as we wanted to better detect the difference in the three test conditions, we added five letters rather than four, each of which represented 0.02 logMAR (see [Figure 1B](#)). The directions of E were generated randomly each time. The spaces between optotypes were fixed at one E size of that line.

Subjects wore polarized glasses coupled with their own best optical correction in all testing conditions when they viewed the dichoptic visual acuity chart. The viewing distance was set to 4 m. Detailed test conditions are shown in [Figure 1C](#). In the dichoptic-optotypes condition, the left eye was shown with the top line of Es in one plane, and the right eye was shown with the bottom line of Es in another plane; these configurations allowed the stimuli to be fused between the eyes. In the monocular condition, the visual acuity of each eye was tested by presenting one line of Es, while the untested eye was occluded with a dark opaque patch. In the dichoptic-luminance condition, the shown visual stimuli were the same as those of the monocular condition for the tested eye; however, the untested eye was presented with a blank screen that had a mean luminance of 96.4 cd/m^2 . Visual acuity was tested by presenting one or two lines of Es in monocular condition; it had no significant difference (see [Supplementary Material 2](#)).

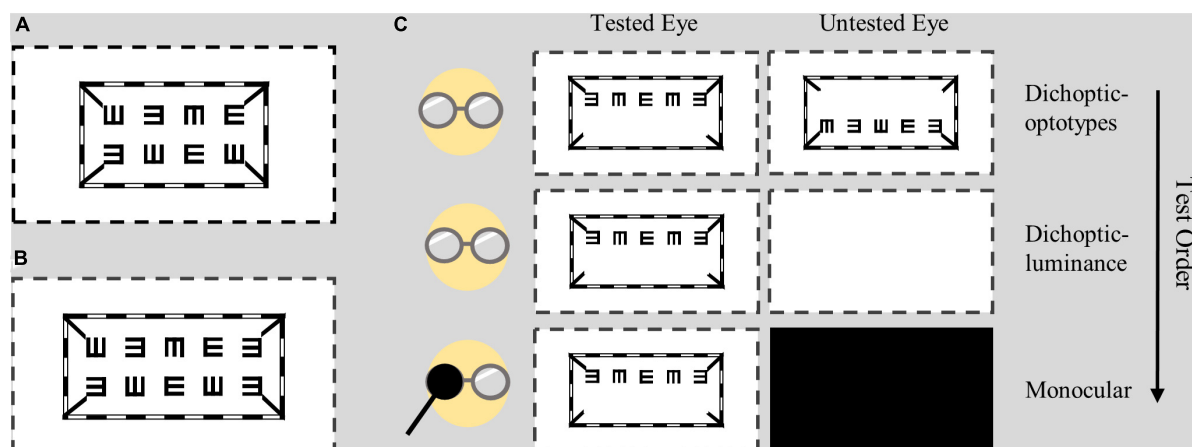


FIGURE 1

Design of the dichoptic visual acuity chart. Two lines of E letters were presented to different eyes, while the envelope and four short lines were presented to both eyes. (A) In Experiment 1, each line had four letters with a letter representing 0.025 logMAR. (B) In Experiment 2, each line had five letters with a letter representing 0.02 logMAR. (C) Test conditions and test order. Subjects wore polarized glasses during all test conditions. The dotted lines indicate the margin of the screen.

Suppression measurement

Binocular phase combination task

We used a binocular phase combination task (Ding and Sperling, 2006; Huang et al., 2009) to quantitatively measure the depth of interocular suppression (Garcia-Perez and Peli, 2019; Min et al., 2021). There were two phases for every trial in this task. First, there was an alignment phase during which dichoptic crosses were shown on the screen. The subjects were asked to align the dichoptic cross into an intact fused cross to ensure that there was a proper fusion between the eyes throughout the experiment. Subsequently, a test phase followed during which two horizontal sign-wave gratings of 1 cycle/degree with a $+22.5^\circ$ or -22.5° phase shift (two configurations) from the center were dichoptically presented to observers through polarized glasses at a distance of 156 cm. Before measuring suppression, we tested for sensory dominance of the normal individuals by fixing the interocular contrast at 1. While measuring suppression, we fixed the contrast of the grating presented to the non-sensory dominant eye at 100% and varied the contrast of the sensory dominant eye at different contrast ratios (i.e., 0, 0.1, 0.2, 0.4, 0.8, and 1). Observers were asked to place the flanking black reference line to the darkest position of the fused sinusoidal grating during each trial. By doing so, we were able to quantitatively measure which eye was more dominant at each trial. Each trial was repeated 8 times. In total, 98 trials [2 (configurations) \times 6 (contrast ratios) \times 8 (repeat)] were conducted in random order. The data of perceived phases were then fitted with an attenuation gain control model (Huang et al., 2009) to obtain the interocular contrast ratio, which is where both eyes contribute equally to binocular vision. The smaller the ratio is, the larger the imbalance between eyes.

Worth 4-dot test

We used a Worth 4-dot (W4D) test to assess suppression in Experiment 2. During the test, subjects were asked to wear red/green anaglyph glasses and were instructed to report the number of dots and then the color of the physical white dot at the bottom at a viewing distance of 4 meters. A W4D score of 0 means no dominance (4 dots with the bottom dot yellow or red and green), 1 means partial suppression (4 dots with the bottom dot red or green), and 2 means strong suppression (only 2 or 3 dots were reported).

Procedure

In Experiment 1, subjects performed the binocular phase combination task. They were able to take a break whenever they wanted to. Then, their visual acuity of each eye was tested using the dichoptic visual acuity chart. The order of the condition was: (1) dichoptic-optotypes condition and (2) monocular condition. The order was not randomized because we wanted to avoid the effect of monocular deprivation (Min et al., 2018); therefore, the dichoptic-optotypes condition was performed before monocular condition. In the dichoptic-optotypes condition, at each logMAR level, subjects were encouraged to report the directions of E from left to right and then from top to bottom (first left eye, then right eye). The test started at the 1.0 logMAR line and was reduced by 0.1 logMAR when the subjects reported the right directions of all 4 optotypes along the one line. If subjects failed to pass the line, the visual acuity of this eye was recorded as 1.0 logMAR minus the corresponding value of right optotypes numbers. Once the test of one eye ended, subjects were only asked to read the lines of the other eye in the subsequent logMAR levels. The performance

of two eyes was recorded. Then, we covered a random eye of subject with a blank opaque patch to measure the monocular visual acuity. After finishing the monocular test of one eye, subjects had 5 min to rest under normal binocular vision. Then the other eye was tested in the monocular condition.

For Experiment 2, the procedure was similar to that used in Experiment 1 except that we measured suppression using a W4D test and inserted a dichoptic-luminance condition between the dichoptic-optotypes and monocular conditions. In the dichoptic-optotypes condition, subjects had to read at least 4 letters in the right directions if they wanted to proceed to the next line. After performing the dichoptic-optotypes condition, we measured one eye's visual acuity of all subjects while their other eye was presented mean luminance. Subsequently, the visual acuity of the other eye was tested in the same viewing condition. On the same day, some subjects performed another testing session of dichoptic-optotypes and monocular conditions after taking a break under normal binocular vision for more than 10 min; this was included in our experimental design so that we could measure the test-retest reliability of the dichoptic letter chart.

Statistical analysis

Data are presented as the mean \pm SD unless otherwise indicated. Statistical analyses were performed using the SPSS 26.0 software package (SPSS Inc., Chicago, IL, United States). A Bland–Altman plot of test-retest reliability was drawn using the GraphPad Prism 8.4.2 software package (GraphPad Software Inc., San Diego, CA, United States). The normal distribution of data was assessed via the Shapiro–Wilk test, and the homogeneity of variance assumption was examined via the Levene test. The baseline comparability of ages in the three groups was compared using one-way analysis of variance (ANOVA). The visual acuity tested under dichoptic conditions and monocular conditions was compared using a paired *t*-test or a paired Wilcoxon rank sum test. The visual acuity change of NDE in the three groups was compared using one-way ANOVA and Fisher's LSD *post hoc* test. Correlation relationships were computed using a Pearson correlation test or the Spearman correlation test, depending on whether the data of interest were normally or non-normally distributed, respectively. $P < 0.05$ was deemed as significant.

Results

Visual acuity in different conditions

Figure 2 shows a plot of visual acuity for each eye in different test conditions. It shows whether there is a relationship (i.e., correlation) between visual acuity data in

monocular and dichoptic conditions. To illustrate, **Figure 2A** shows asymmetric CVA of two eyes in treated and untreated individuals with amblyopia. The visual acuity of DE between these two conditions was not significantly different in the amblyopia (0.02 ± 0.08 vs. 0.01 ± 0.08 ; $t = -1.580$, $P = 0.131$) and treated groups (0.02 ± 0.04 vs. 0.01 ± 0.04 ; $Z = -1.311$, $P = 0.190$). However, the visual acuity of NDE in the amblyopia group was significantly reduced to 0.45 ± 0.15 logMAR in the dichoptic-optotypes condition from 0.28 ± 0.18 logMAR in the monocular condition ($t = -9.067$, $P < 0.001$). For the treated group, the visual acuity of NDE in the amblyopia group was significantly reduced to 0.15 ± 0.10 logMAR in the dichoptic-optotypes condition from 0.08 ± 0.04 logMAR in the monocular condition ($t = -4.162$, $P = 0.001$). There was no difference between these two conditions in the normal observers (NDE: $t = -1.739$, $P = 0.104$; DE: $t = -0.653$, $P = 0.524$).

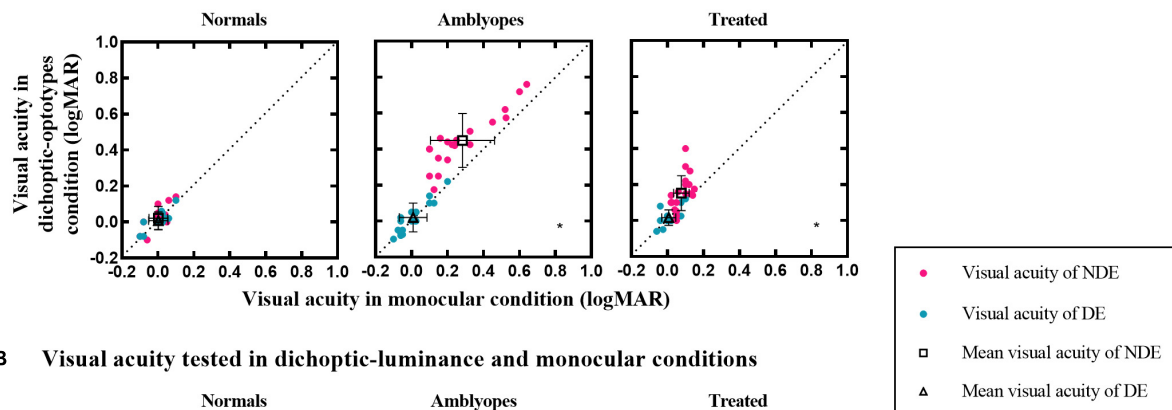
To examine whether the visual acuity change in the dichoptic-optotypes condition was caused by the mean luminance of the fixing region of the untested eye, we compared the visual acuity results between the dichoptic-luminance condition and monocular condition (see **Figure 2B**). Visual acuity did not change in dichoptic-luminance and monocular conditions for each eye of the three groups (Normal group: $Z = -1.289$, $P = 0.197$ for NDE and $Z = -0.816$, $P = 0.414$ for DE; Amblyopia group: $t = 0.414$, $P = 0.691$ for NDE and $Z = -0.365$, $P = 0.715$ for DE; Treated group: $t = -1.246$, $P = 0.244$ for NDE and $t = 0.287$, $P = 0.780$ for DE).

The visual acuity change of non-dominant eye

CVA of NDE in the dichoptic-optotypes condition and monocular condition (CVA = visual acuity of NDE in the dichoptic-optotypes condition—visual acuity of NDE in the monocular condition) were computed as an index to represent the impact of suppression on visual acuity. As shown in **Figure 3A**, CVA in three groups was significantly different from each other ($F = 18.118$, $P < 0.001$). The CVA of the amblyopia group was 0.16 ± 0.08 logMAR, which was more than 0.08 ± 0.08 logMAR in the treated group and 0.02 ± 0.04 logMAR in the normal group (both $P < 0.001$ in *post hoc* test). A significant difference was also found between the treated and normal groups ($P = 0.026$ in *post hoc* test). The ages of the three groups were not significantly different ($F = 1.093$, $P = 0.343$).

Figure 3B shows plots of the CVA as a function of the depth of suppression tested; the data were obtained using the binocular phase combination task and W4D test. A significant negative correlation was found between the CVA and the interocular contrast ratio at the balance point (Spearman's $\rho = -0.72$, $P = 0.005$). This result indicates that there is a positive correlation between the CVA and the depth of suppression. However, no significant relationship was found between CVA

A Visual acuity tested in dichoptic-optotypes and monocular conditions



B Visual acuity tested in dichoptic-luminance and monocular conditions

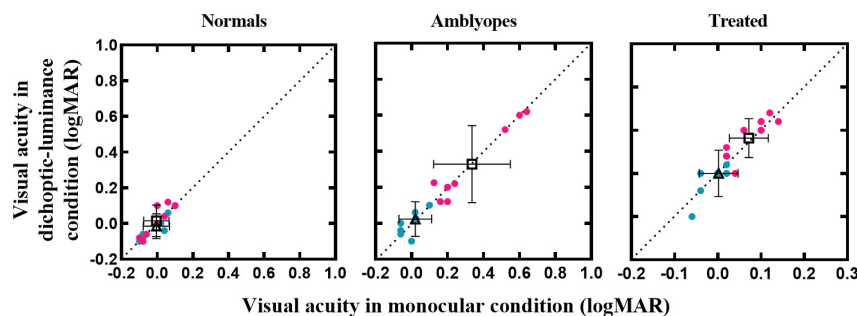


FIGURE 2

Visual acuity of the three groups in different test conditions. (A) Visual acuity under dichoptic-optotypes condition as a function of visual acuity under monocular condition. (B) Visual acuity under dichoptic-luminance condition as a function of visual acuity under monocular condition. The black dotted lines indicate the same visual acuity under two conditions. Black hollow squares and black hollow triangles represent mean visual acuity of the NDE and DE, respectively. Error bars denote the range of \pm SD. A black asterisk is on the lower right of the panel which shows significant difference between two test conditions.

and suppression tested by the W4D test (Spearman's $\rho = 0.380$, $P = 0.067$).

Test-retest reliability

We retested the visual acuity in dichoptic-optotypes and monocular conditions in 18 subjects (5 normal, 6 amblyopes, and 7 treated) to assess the test-retest reliability of the novel dichoptic visual acuity test in different groups. Figure 4A shows the plot of the visual acuity of each eye of the second test against the visual acuity measured from the first test. The correlation coefficients were greater than 0.90 for every group under each condition. The correlation coefficients were 0.91 ($P = 0.0003$) in dichoptic-optotypes condition and 0.96 ($P < 0.0001$) in monocular condition for normal group, 0.99 and 0.98 (both $P < 0.0001$) for amblyopia group, and 0.95 and 0.97 (both $P < 0.0001$) for treated group. Figure 4B shows a Bland-Altman difference plot of two measurements. The differences (first test-second test) are plotted against the mean values for each subject. The mean differences between the first and second tests and the 95% confidence interval (CI) limits of agreement of the normal group were 0.02 (95% CI, -0.04 to 0.08) in

dichoptic-optotypes condition and 0.00 (95% CI, -0.04 to 0.04) in monocular condition. Those of the amblyopic group were 0.03 (95% CI, -0.06 to 0.12) and -0.01 (95% CI, -0.09 to 0.08). Those of the treated group were 0.00 (95% CI, -0.08 to 0.09) and 0.00 (95% CI, -0.03 to 0.04), respectively. The proportion of visual acuity difference between two measures that fell within less than 1 line (0.1 logMAR) was 93%, suggesting a robust test-retest reliability.

Discussion

In this study, using the novel dichoptic visual acuity chart that exhibits a robust test-retest reliability, we found that the visual acuity of the amblyopic eye and the previous amblyopic eye decreased when the other eye (i.e., dominant eye) viewed optotypes but not mean luminance. Also, the magnitude of decreased visual acuity was correlated with the depth of interocular suppression.

We confirmed what has already been shown in previous studies: the visual acuity of the amblyopic eye was reduced in the dichoptic condition. The maximal extent of reduction was less than that in the study of von Noorden and Leffler (1966) in

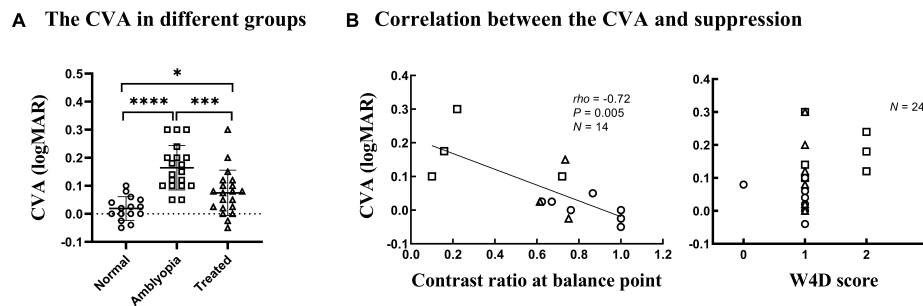


FIGURE 3

CVA in the three groups and correlation with suppression. Each dot represents the results of one patient. (A) The CVA in three groups. **** $P < 0.0001$; *** $P < 0.001$; * $P < 0.05$. (B) CVA as a function of the contrast ratio at the balance point tested by the binocular phase combination task in Experiment 1 and the W4D scores tested in Experiment 2. The greater the contrast ratio at the balance point is close to 1, the more balanced or less suppressed.

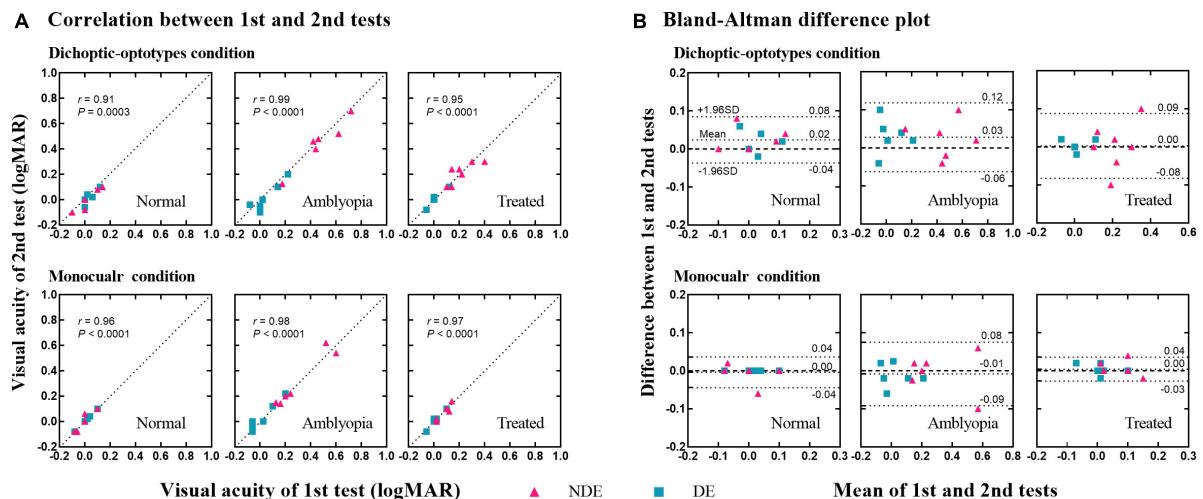


FIGURE 4

Test-retest reliability of the dichoptic visual acuity chart. (A) Correlation between the 1st and 2nd tests. The dotted lines indicate the line of equality (1st test = 2nd test). (B) Difference in visual acuity between the 1st and 2nd tests (1st test–2nd test) as a function of the mean value of the two tests [(1st test + 2nd test)/2]. Three dotted lines indicate the 95% upper limits of agreement, the bias, and the 95% lower limits of agreement, respectively.

which the difference in visual acuity in the monocular condition and dichoptic condition ranged from 0 to 5 lines; however, there were 0–3 lines in this study. This may be because some of the amblyopic participants in the study of Narasimhan et al. (2012) had strabismus, which could exhibit more severe suppression than anisometropic amblyopia without strabismus.

Measurement of visual acuity of the amblyopic eye has been the standard way to diagnose and track the recovery of amblyopia in the clinic (Mintz-Hittner and Fernandez, 2000; Stewart et al., 2003). However, this approach is strictly monocular and does not reflect the binocular mechanism. Furthermore, amblyopia could be due to both monocular attenuation of the amblyopic eye and imbalanced interocular suppression—a binocular process—from the fellow eye to the amblyopic eye (Huang et al., 2011). Our results indicate that

visual acuity of the amblyopic eye or the previous amblyopic eye was different depending on the visual state of the other eye. For instance, when the amblyopic eye was tested while the fellow was occluded with an opaque patch, the visual acuity was seemingly intact, thereby indicating that suppression was minimized. However, we found that the visual acuity of the amblyopic eye was worse when the fellow eye was open. The visual acuity change tested in our study reflects the binocular mechanism because our results showed that this residual suppression in treated amblyopia has functional significance in visual acuity.

Our results indicate that the dichoptic letter chart has a good test-retest reliability. Although this tool does not measure the depth of suppression directly, it can quantitatively show the visual acuity change caused by suppression, which is correlated to its depth. Furthermore, the dichoptic visual

acuity chart used here is intuitive and easy to complete for children. In future, we suggest that investigators consider using the change in visual acuity (as measured with the dichoptic visual acuity chart) between dichoptic-optotypes condition and monocular condition as an outcome measurement so that they could better show the treatment efficacy of amblyopia.

As the fellow eye was fixed at a region of mean luminance while the amblyopic eye was fixed at optotypes in this study, we tested whether the visual acuity change was caused by mean luminance. We have the same result as Vedomurthy et al. (2007) and Zhou et al. (2014) that mean luminance in the fellow eye does not affect the visibility of the amblyopic eye. This confirms that the reduced visual acuity observed in our study might not be caused by mean luminance. In contrast, Jia et al. (2015) found that the mean luminance stimulus of the fellow eye reduced the contrast sensitivity of the amblyopic eye. Lai et al. (2011) also found that the visual acuity and contrast sensitivity of the amblyopic eye were reduced when the fellow eye was partially patched (by presenting square mean luminance to the fellow eye). Several factors may account for these discrepancies: (1) The measuring targets: optotypes were used as measuring targets of the tested eye in our study and those of Vedomurthy et al. (2007) and Zhou et al. (2014). These optotypes had different bandwidths from sine-gratings used in the study of Jia et al. (2015); (2) interocular contrast difference: 78% contrast at the edge of square mean luminance in the fellow eye and 20% Weber contrast optotypes in the amblyopic eye were used in the study of Lai et al. (2011). Presenting optotypes to the untested eye rather than mean luminance could be more useful if one was interested in measuring the visual acuity change of the tested eye caused by suppression. This would be because two eyes would be presented with the same luminance and contrast level.

Limitations

Our novel visual acuity chart could be limited by the possible influence of the crowding effect (Stuart and Burian, 1962; Levi, 2008). The crowding effect can be induced when the flankers and target are presented to different eyes (Flom et al., 1963). For example, the presentation of two lines of Es in dichoptic-optotypes condition and one line of Es in monocular condition could produce different crowding effects in amblyopia. This could also reduce the tested visual acuity under dichoptic-optotypes condition. If the crowding effect impacts the outcome, the visual acuity that was tested using two lines of Es should be worse than that tested using one line of Es. We conducted a control experiment (see [Supplementary Material](#)) and found no significant difference in visual acuity when it was tested using one line and two lines of Es. Thus, the crowding effect may not influence our conclusions. The

crowding effect is found more in amblyopia with strabismus (Hess et al., 2001; Hariharan et al., 2005) but less in pure anisometropic amblyopia (Bonneh et al., 2004; Greenwood et al., 2012). In this study, our patients were, or used to be anisometropic amblyopia, implying that the crowding effect could be quite minimal.

If one is interested in using the dichoptic visual acuity chart to examine amblyopes with strabismus, one should account for the crowding by presenting two lines of Es to the tested eye under monocular condition.

Conclusion

In conclusion, our study suggests that a dichoptic visual acuity chart with optotypes presented to the untested eye can be a feasible and reliable option to measure the impact of suppression on visual acuity in both the laboratory and the clinic for studying and treating amblyopia.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee on Biomedical Research of West China Hospital of Sichuan University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

BZ contributed to the study design, data acquisition, analysis, interpretation, and manuscript writing and revision. ML contributed to the study design, experimental equipment, data acquisition, interpretation, and manuscript writing and revision. LL contributed to the study design, experimental equipment, data interpretation, and approval of the final version for publication. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (82070996) and the Technical Innovation Research and Development Project of Science and Technology Bureau in Chengdu (2021-YF05-00670-SN).

Acknowledgments

We thank the subjects and families for their kind participation in our research. We also thank Seung Hyun (Sam) Min (www.ses21.com) for his very kind help in correcting English.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships

that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2022.860620/full#supplementary-material>

References

- Birch, E. E., Jost, R. M., De La Cruz, A., Kelly, K. R., Beauchamp, C. L., Dao, L., et al. (2019). Binocular amblyopia treatment with contrast-rebalanced movies. *J. AAPOS*. 160, e1–160.e5. doi: 10.1016/j.jaapos.2019.02.007
- Birch, E. E., Morale, S. E., Jost, R. M., De La Cruz, A., Kelly, K. R., Wang, Y. Z., et al. (2016). Assessing Suppression in Amblyopic Children With a Dichoptic Eye Chart. *Invest. Ophthalmol. Vis. Sci.* 57, 5649–5654. doi: 10.1167/jovs.16-19986
- Bonneh, Y. S., Sagi, D., and Polat, U. (2004). Local and non-local deficits in amblyopia: acuity and spatial interactions. *Vision Res.* 44, 3099–3110. doi: 10.1016/j.visres.2004.07.031
- Chen, H., He, Z., Xu, J., Mao, Y., Liang, Y., Lin, D., et al. (2019). A Convenient and Robust Test to Quantify Interocular Suppression for Children With Amblyopia. *Iperception* 10:2041669519864971. doi: 10.1177/2041669519864971
- Chen, S., Min, S. H., Cheng, Z., Xiong, Y., Yu, X., Wei, L., et al. (2021). Binocular visual deficits at mid to high spatial frequency in treated amblyopes. *iScience* 24:102727. doi: 10.1016/j.isci.2021.102727
- Chen, Y., Wang, J., Shi, H., Wang, X., and Feng, L. (2017). Sensory Eye Dominance in Treated Anisometropic Amblyopia. *Neural Plast.* 2017:9438072. doi: 10.1155/2017/9438072
- Coren, S., and Kaplan, C. P. (1973). Patterns of ocular dominance. *Am. J. Optom. Arch. Am. Acad. Optom.* 50, 283–92.
- Ding, J., and Sperling, G. (2006). A gain-control theory of binocular combination. *Proc. Natl. Acad. Sci. U.S.A.* 103, 1141–1146. doi: 10.1073/pnas.0509629103
- Flom, M. C., Weymouth, F. W., and Kahneman, D. (1963). Visual Resolution and Contour Interaction. *J. Opt. Soc. Am.* 53, 1026–32. doi: 10.1364/josa.53.001026
- Fu, Z., Hong, H., Su, Z., Lou, B., Pan, C. W., and Liu, H. (2019). Global prevalence of amblyopia and disease burden projections through 2040: a systematic review and meta-analysis. *Br. J. Ophthalmol.* 104, 1164–1170. doi: 10.1136/bjophthalmol-2019-314759
- Garcia-Perez, M. A., and Peli, E. (2019). Psychophysical Tests Do Not Identify Ocular Dominance Consistently. *Iperception* 10:2041669519841397. doi: 10.1177/2041669519841397
- Greenwood, J. A., Tailor, V. K., Sloper, J. J., Simmers, A. J., Bex, P. J., and Dakin, S. C. (2012). Visual acuity, crowding, and stereo-vision are linked in children with and without amblyopia. *Invest. Ophthalmol. Vis. Sci.* 53, 7655–65. doi: 10.1167/jovs.12-10313
- Hariharan, S., Levi, D. M., and Klein, S. A. (2005). "Crowding" in normal and amblyopic vision assessed with Gaussian and Gabor C's. *Vision Res.* 45, 617–633. doi: 10.1016/j.visres.2004.09.035
- Hess, R. F., Dakin, S. C., Tewfik, M., and Brown, B. (2001). Contour interaction in amblyopia: scale selection. *Vision Res.* 41, 2285–2296. doi: 10.1016/s0042-6989(01)00099-2
- Hess, R. F., and Thompson, B. (2015). Amblyopia and the binocular approach to its therapy. *Vision Res.* 114, 4–16. doi: 10.1016/j.visres.2015.02.009
- Hess, R. F., Thompson, B., and Baker, D. H. (2014). Binocular vision in amblyopia: structure, suppression and plasticity. *Ophthalmic. Physiol. Opt.* 34, 146–162. doi: 10.1111/opo.12123
- Huang, C. B., Zhou, J., Lu, Z. L., Feng, L., and Zhou, Y. (2009). Binocular combination in anisometropic amblyopia. *J. Vis.* 9:17. doi: 10.1167/9.3.17
- Huang, C. B., Zhou, J., Lu, Z. L., and Zhou, Y. (2011). Deficient binocular combination reveals mechanisms of anisometropic amblyopia: signal attenuation and interocular inhibition. *J. Vis.* 11:4. doi: 10.1167/11.6.4
- Huang, P. C., Baker, D. H., and Hess, R. F. (2012). Interocular suppression in normal and amblyopic vision: spatio-temporal properties. *J. Vis.* 12:29. doi: 10.1167/12.11.29
- Jia, W., Lan, F., Zhao, X., Lu, Z. L., Huang, C. B., Zhao, W., et al. (2018). The effects of monocular training on binocular functions in anisometropic amblyopia. *Vision Res.* 152, 74–83. doi: 10.1016/j.visres.2017.02.008
- Jia, W., Zhou, J., Lu, Z. L., Lesmes, L. A., and Huang, C. B. (2015). Discriminating anisometropic amblyopia from myopia based on interocular inhibition. *Vision Res.* 114, 135–141. doi: 10.1016/j.visres.2015.02.003
- Kehrein, S., Kohnen, T., and Fronius, M. (2016). Dynamics of Interocular Suppression in Amblyopic Children during Electronically Monitored Occlusion Therapy: first Insight. *Strabismus* 24, 51–62. doi: 10.3109/09273972.2016.1170047
- Kwon, M., Wiecek, E., Dakin, S. C., and Bex, P. J. (2015). Spatial-frequency dependent binocular imbalance in amblyopia. *Sci. Rep.* 5:17181. doi: 10.1038/srep17181
- Lai, X. J., Alexander, J., He, M., Yang, Z., and Suttle, C. (2011). Visual functions and interocular interactions in anisometropic children with and without amblyopia. *Invest. Ophthalmol. Vis. Sci.* 52, 6849–6859. doi: 10.1167/jovs.10-6755

- Lai, X. J., Alexander, J., He, M. G., Yang, Z. K., and Suttle, C. (2012). A novel apparatus for interocular interaction evaluation in children with and without anisometropic amblyopia. *Clin. Exp. Optom.* 95, 410–420. doi: 10.1111/j.1444-0938.2012.00753.x
- Levi, D. M. (2008). Crowding - an essential bottleneck for object recognition: a minireview. *Vision Res.* 48, 635–654. doi: 10.1016/j.visres.2007.12.009
- Li, J., Thompson, B., Lam, C. S., Deng, D., Chan, L. Y., Machara, G., et al. (2011). The role of suppression in amblyopia. *Invest. Ophthalmol. Vis. Sci.* 52, 4169–4176. doi: 10.1167/iovs.11-7233
- Mao, Y., Min, S. H., Chen, S., Gong, L., Chen, H., Hess, R. F., et al. (2020). Binocular Imbalance in Amblyopia Depends on Spatial Frequency in Binocular Combination. *Invest. Ophthalmol. Vis. Sci.* 61:7. doi: 10.1167/iovs.61.8.7
- Min, S. H., Baldwin, A. S., Reynaud, A., and Hess, R. F. (2018). The shift in ocular dominance from short-term monocular deprivation exhibits no dependence on duration of deprivation. *Sci. Rep.* 8:17083. doi: 10.1038/s41598-018-35084-1
- Min, S. H., Gong, L., Baldwin, A. S., Reynaud, A., He, Z., Zhou, J., et al. (2021). Some psychophysical tasks measure ocular dominance plasticity more reliably than others. *J. Vis.* 21:20. doi: 10.1167/jov.21.8.20
- Mintz-Hittner, H. A., and Fernandez, K. M. (2000). Successful amblyopia therapy initiated after age 7 years: compliance cures. *Arch. Ophthalmol.* 118, 1535–41. doi: 10.1001/archophth.118.11.1535
- Multi-Ethnic Pediatric Eye Disease, and Study, G. (2009). Prevalence and causes of visual impairment in African-American and Hispanic preschool children: the multi-ethnic pediatric eye disease study. *Ophthalmology* 116, 1990–2000.e1. doi: 10.1016/j.ophtha.2009.03.027
- Multi-ethnic Pediatric Eye Disease Study, G. (2008). Prevalence of amblyopia and strabismus in African American and Hispanic children ages 6 to 72 months the multi-ethnic pediatric eye disease study. *Ophthalmology* 115, 1229–1236.e1. doi: 10.1016/j.ophtha.2007.08.001
- Narasimhan, S., Harrison, E. R., and Giaschi, D. E. (2012). Quantitative measurement of interocular suppression in children with amblyopia. *Vision Res.* 66, 1–10. doi: 10.1016/j.visres.2012.06.007
- Pugh, M. (1954). Foveal vision in amblyopia. *Br. J. Ophthalmol.* 38, 321–331. doi: 10.1136/bjo.38.6.321
- Stewart, C. E., Moseley, M. J., and Fielder, A. R. (2003). Defining and measuring treatment outcome in unilateral amblyopia. *Br. J. Ophthalmol.* 87, 1229–1231. doi: 10.1136/bjo.87.10.1229
- Stuart, J. A., and Burian, H. M. (1962). A study of separation difficulty: its relationship to visual acuity in normal and amblyopic eyes. *Am. J. Ophthalmol.* 53, 471–477. doi: 10.1016/0002-9394(62)94878-x
- Vedamurthy, I., Suttle, C. M., Alexander, J., and Asper, L. J. (2007). Interocular interactions during acuity measurement in children and adults, and in adults with amblyopia. *Vision Res.* 47, 179–188. doi: 10.1016/j.visres.2006.08.017
- von Noorden, G. K., and Leffler, M. B. (1966). Visual acuity in strabismic amblyopia under monocular and binocular conditions. *Arch. Ophthalmol.* 76, 172–7. doi: 10.1001/archophth.1966.03850010174004
- Wallace, D. K., Repka, M. X., Lee, K. A., Melia, M., Christiansen, S. P., Morse, C. L., et al. (2018). Amblyopia Preferred Practice Pattern(R). *Ophthalmology* 125, 105–142. doi: 10.1016/j.ophtha.2017.10.008
- Williams, C., Northstone, K., Howard, M., Harvey, I., Harrad, R. A., and Sparrow, J. M. (2008). Prevalence and risk factors for common vision problems in children: data from the ALSPAC study. *Br. J. Ophthalmol.* 92, 959–964. doi: 10.1136/bjo.2007.134700
- Zhao, W., Jia, W. L., Chen, G., Luo, Y., Lin, B., He, Q., et al. (2017). A complete investigation of monocular and binocular functions in clinically treated amblyopia. *Sci. Rep.* 7:10682. doi: 10.1038/s41598-017-11124-0
- Zhou, J., Jia, W., Huang, C. B., and Hess, R. F. (2013). The effect of unilateral mean luminance on binocular combination in normal and amblyopic vision. *Sci. Rep.* 3:2012. doi: 10.1038/srep02012
- Zhou, J., McNeal, S., Babu, R. J., Baker, D. H., Bobier, W. R., and Hess, R. F. (2014). Time course of dichoptic masking in normals and suppression in amblyopes. *Invest. Ophthalmol. Vis. Sci.* 55, 4098–4104. doi: 10.1167/iovs.14-13969
- Zhou, J., Reynaud, A., Yao, Z., Liu, R., Feng, L., Zhou, Y., et al. (2018). Amblyopic Suppression: passive Attenuation, Enhanced Dichoptic Masking by the Fellow Eye or Reduced Dichoptic Masking by the Amblyopic Eye? *Invest. Ophthalmol. Vis. Sci.* 59, 4190–4197. doi: 10.1167/iovs.18-24206



Assessing Heterogeneity in Students' Visual Judgment: Model-Based Partitioning of Image Rankings

Miles Tallon^{1,2*}, Mark W. Greenlee¹, Ernst Wagner³, Katrin Rakoczy⁴, Wolfgang Wiedermann⁵ and Ulrich Frick²

¹Department of Experimental Psychology, University of Regensburg, Regensburg, Germany, ²HSD Research Centre Cologne, HSD University of Applied Sciences, Cologne, Germany, ³Academy of Fine Art Munich, Munich, Germany, ⁴Institute for School Education and Empirical Educational Research, Justus-Liebig University, Gießen, Germany, ⁵Missouri Prevention Science Institute and Department of Educational, School, and Counseling Psychology, University of Missouri, Columbia, KY, United States

OPEN ACCESS

Edited by:

Alyssa A. Brewer,
University of California,
Irvine, United States

Reviewed by:

Amadou Sawadogo,
Félix Houphouët-Boigny University,
Côte d'Ivoire
Katherine Rebecca Storrs,
Justus Liebig University Giessen,
Germany

*Correspondence:

Miles Tallon
m.tallon@hs-doeper.de

Specialty section:

This article was submitted to
Perception Science,
a section of the journal
Frontiers in Psychology

Received: 22 February 2022

Accepted: 20 June 2022

Published: 10 August 2022

Citation:

Tallon M, Greenlee MW, Wagner E,
Rakoczy K, Wiedermann W and
Frick U (2022) Assessing
Heterogeneity in Students' Visual
Judgment: Model-Based Partitioning
of Image Rankings.
Front. Psychol. 13:881558.
doi: 10.3389/fpsyg.2022.881558

Differences in the ability of students to judge images can be assessed by analyzing the individual preference order (ranking) of images. To gain insights into potential heterogeneity in judgement of visual abstraction among students, we combine Bradley–Terry preference modeling and model-based recursive partitioning. In an experiment a sample of 1,020 high-school students ranked five sets of images, three of which with respect to their level of visual abstraction. Additionally, 24 art experts and 25 novices were given the same task, while their eye movements were recorded. Results show that time spent on the task, the students' age, and self-reported interest in visual puzzles had significant influence on rankings. Fixation time of experts and novices revealed that both groups paid more attention to ambiguous images. The presented approach makes the underlying latent scale of visual judgments quantifiable.

Keywords: visual abstraction, assessment, Bradley–Terry model, model-based partitioning, ranking, art education, visual literacy

INTRODUCTION

This study is part of a larger research project on the assessment of Visual Literacy (VL) and how VL can be fostered in art education (Frick et al., 2020). VL, a core competency in art education, comprises the ability to evaluate artwork with respect to aesthetic value. The Common European Framework of Reference for Visual Literacy (CEFR-VL; Wagner and Schönau, 2016) defines *judging* (or evaluating) images as the ability to formulate a justified statement or estimation about images and artistic creations. We define visual abstraction as a prerequisite for aesthetic judgment and as a latent variable in a visual judgement task. The method described here contributes to determine essential variables that impact the judgment of latent image features (exemplified by visual abstraction) and in return might help teachers detect and promote students' development of artistic skills. Furthermore, identifying critical variables that influence students' visual judgments may be important for empirical art education research. The aim of the present study is to investigate students' ability, on the one hand, as well as that of experts and novices, on the other hand, to judge images based on the level of perceived visual abstraction, while placing a focus on the identification of biographical and psychological

characteristics that influence these judgments. Aesthetic judgments are not only influenced by the properties of the items being judged but they are influenced by additional factors such as expertise and personal experience (Child, 1965; Nodine et al., 1993; Jacobsen, 2004; Hayn-Leichsenring et al., 2020; McCormack et al., 2021). For example, Chamorro-Premuzic and Furnham (2004) showed that university students with higher interests in art tend to score higher on art judgment tasks and that these judgments were significantly related to both personality and intelligence.

Every artwork, whether figurative or not, is a form of abstraction (Witkin, 1983; Gortais, 2003). However, the measurement of the perceived level of visual abstraction in artworks remains challenging. A study that specifically tried to measure the perceived level of visual abstraction used visual analog scales to rate artworks as “abstract” and found contrast effects due to sequential presentation of high vs. low abstract paintings on participants’ judgments (Specht, 2007). Other studies explored the preference judgment of abstract art measured by Likert-scale ratings and revealed a preference for the artists’ original compositions (McManus et al., 1993; Furnham and Rao, 2002). Efforts to quantify visual abstraction in artworks were also made by Chatterjee et al. (2010): their Assessment of Art Attributes instrument (AAA) includes “abstraction” as a conceptual-representational attribute. The level of abstraction is measured *via* a Likert-scale rating and training slides with example images as anchors. Another assessment tool, the Rating Instrument for Two-Dimensional Pictorial Works (RizbA; Schoch and Ostermann, 2020), consists of 26 six-point Likert-scale items, including two questions regarding the mode of concrete and abstract representation.

However, when underlying image features are latent (e.g., the extent to which a given image is abstract) metric scales may fall short when asked to judge these items by, for example, assigning a number from 1 to 10. Typical disadvantages of the use of such absolute measures may include anchor effects (Furnham and Boo, 2011) and end-aversion bias (Streiner and Norman, 2008) among others (Choi and Pak, 2005). It is often easier to compare items to each other, e.g., in a series of paired comparison (PC) tasks. Such comparative measures can be analyzed with Bradley–Terry (BT) models (Bradley and Terry, 1952), also referred to as Bradley–Terry–Luce models. BT models are a popular method to uncover a latent preference scale of objects/items from paired comparison data (Cattelan, 2012). For example, BT models are frequently used to determine the best sport teams (Cattelan et al., 2013), to analyze consumer-specific preferences (Dittrich et al., 2000), or to determine the perceived harm of psychotropic substances (Wiedermann et al., 2014). When multiple objects (images) are compared simultaneously, ranking tasks (e.g., ranking images according to their level of abstraction) constitute valuable alternatives to PCs. Ranking data can then be transformed into derived PC patterns (Francis et al., 2010).

The present study focuses on potential heterogeneity in visual judgments. Potential differences in visual judgments were evaluated in two samples: a sample of high-school students and an additional sample comprising art experts (art educators,

artists, designers) and novices (art laypersons). In the student sample, self-reported visual skills and demographic variables are used to detect potential differences in students’ performance to rank different sets of images based on their level of visual abstraction. In the experts and novices sample eye movements were additionally recorded during the image ranking task. Eye movement indicators are used to analyze the distribution of attention (Jarodzka et al., 2017; Brams et al., 2019). Eye tracking, in particular as an exploratory tool, can enhance the multidisciplinary field of VL research, as it visualizes cognitive processes involved in visual problem solving and art perception (Brumberger, 2021). Visualizing the solution process with VL-experts’ and novices’ eye-movements can be used to uncover cognitive processes that differ between the expert and novice groups and may further reveal difficult or ambiguous image sets.

This study uses model-based partitioning as a method to analyze what underlies the variability in visual judgments. We use a recently published approach that combines Bradley–Terry (BT) models with model-based recursive partitioning (trees) to detect preference heterogeneity in subgroups (Wiedermann et al., 2021). BT models are well-suited for (art) educational assessment tasks, in which students are instructed to rank images based on given criteria. From a methodological perspective the use of BT models in combination with recursive partitioning is studied for its potential when applied to art education assessment. The reason for this is that conventional statistical analysis of interaction effects may fall short when tasked to address the complex moderation processes of visual judgments. The method used here enables researchers to differentiate between the effects of student characteristics and learning interventions on latent preference rankings more closely. The study addresses the following research questions: What effects do self-reported visual skills and student characteristics have on the order of images when they are ranked according to visual abstraction? Do VL-experts and novices differ in their ranking patterns and solution strategies?

MATERIALS AND METHODS

Subjects and Stimuli

Sample I comprised 1,020 students of which 987 worked on the ranking tasks and filled out the questionnaire. A total of 52 classes (9th to 13th grade) from 29 schools in Germany took part in the study. Two classes did not receive the questionnaire and one class could not be offered the ranking task due to technical difficulties. To control for potentially nested effects of classrooms, intraclass correlation coefficients (ICCs) for intended rankings were calculated on each image set. Due to low values (ICCs range from 0.01 to 0.03, for calculations see Chakraborty and Sen, 2016), no multi-level adjustments were necessary. Overall, 52% of participants were female, the average age was 15.34 years ($SD=2.96$). Schools were recruited in the federal states of Hessen, North-Rhine Westphalia, Schleswig-Holstein, and Rhineland Palatinate *via* leaflets, letters and recommendations. Data collection was conducted in classrooms with up to 30 students ($M=20.8$,

$SD=5.10$). The image ranking task was part of a VL assessment test battery, including demographic questions, art grade, and the following questions regarding artistic ability and self-perceived art skills (S1–S5):

- If you had to rank all of your classmates according to their abilities in the subject of art, where would you rank yourself? [S1; scored 1 (as one of the worst) to 5 (as one of the best)]
- How good are you at art in general? [S2; scored 1 (very bad) to 5 (very good)]
- How good are you in theoretical content (art theory; e.g. interpreting pictures, understanding art history)? [S3; scored 1 (very bad) to 5 (very good)]
- How good are you in practical activities in art class (e.g., painting, drawing, drafting, and designing)? [S4; scored 1 (very bad) to 5 (very good)]
- Compared to your skills in other school subjects: How well do you rate your art skills? [S5; scored 1 (much worse) to 5 (much better)]

Additionally the following self-reported visual skills were rated on a scale from 1 (strongly disagree) to 4 (strongly agree): Photographic memory (PM): “I have a ‘photographic memory’”; Spatial orientation (SO): “When I see a photograph of a geometric object, I can imagine what it looks like from behind”; Long-term memory (LM): “I can remember small details in pictures”; Imagination (IM): “I can easily picture things mentally”; and Interest in visual puzzles (IP): “I like to solve picture puzzles.”

Sample II comprised 51 participants of which 49 participants had qualitatively sufficient eye-tracking data to be included for further analyses. Experts and novices were screened based on their experience and interest or profession in the visual arts. The expert group ($n=24$) consisted of photographers, artists, designers, and art students. The novice group ($n=25$) consisted of students and adults from various educational institutions who were not associated with academic or professional work in the visual arts. The mean age of participants were $M=29.08$ years ($SD=12.55$). The participants in sample II were assessed individually in seminar or laboratory rooms (e.g., at the Academy of Fine Arts in Munich).

In sample I school classes were offered a lump sum of 100€ as collective compensation. In sample II student participants each received 20€ as compensation. Participants from the expert group, who were generally interested in the subject of visual literacy and eye tracking, took part without further incentive. All participants and their legal representatives, respectively, gave written consent before participating in this study. The study was approved by the Ethics Committee of Research of the Leibniz Institute for Research and Information in Education, Frankfurt am Main (DIPF, 01JK1606A).

Ranking Task

We used images with varying level of visual abstraction, i.e., image sets that represent the gradual process of transforming figurative artwork to non-figurative artwork (Viola et al., 2020). As every work of art uses some level of abstraction, many

artworks could be investigated. Therefore images were curated (or created) by visual arts professionals from the board of the European Network for Visual Literacy (ENViL). Image sets were chosen based on the likelihood of being discussed in art class, representing a varying degree of abstraction.

Overall, five ranking tasks were presented on Android tablets with 10.1 inch screen size (Andrews et al., 2018). Subjects ranked five images, resulting in a total of $\binom{5}{2}=10$ paired comparisons for each set of images (with a total of $5!=120$ possible combinations; see **Table 1**). All participants were presented with the same initial ordering of images and were instructed to rank each image according to two characteristics presented below each image set. The image sets included:

1. geometric figures
2. dogs
3. bull images, inspired by Pablo Picasso's *Bull* lithographs (MacTaggart, 2021)
4. Mondrian trees
5. salt packages (only presented in sample I)

Images had to be ranked according to the following image characteristics: starting with an introductory item to make sure that participants understood the task (“geometric figures”), from round to edgy, the items “dogs,” “bull images,” and “Mondrian trees” had to be ranked by level of visual abstraction; from most realistic to most abstract. Additionally, as a control condition, perceived expensiveness (from cheap to expensive) of items (“salt packages”) was assessed. In contrast to the evaluation of image abstraction, rankings based on unknown prices should stand out as visible outliers compared to the other rankings. This was used in an attempt to investigate potential uncertainty of judgments and how this variability may affect the BT ranking results on group level. The ordering ($a > b > c > d > e$) of images was consensually decided by VL experts from ENViL. Participants used a touchscreen to select and drop each image into empty slots presented below the images (see **Figure 1**). The image rankings are then analyzed to gain insights into the possible effects of the participant characteristics on the perceived judgment of abstraction.

Eye Tracking

Each participant in sample II wore eye-tracking glasses (SMI ETG 2w Analysis Pro) during task performance. Eye movements were recorded at 60Hz. A 3-point calibration was performed on the tablet for each participant. All participants had normal or corrected to normal eyesight. Fixations were mapped onto corresponding reference images using SMI fixation-by-fixation semantic gaze mapping (Vansteenkiste et al., 2015). Areas of Interest (AOIs) were drawn on each image to assess fixation time and number of fixations spend on each image. Eye-movement events were determined by the SMI velocity-based algorithm (Engbert et al., 2016). Eye-tracking data, i.e., number of fixations, fixation duration and heatmaps were analyzed with SMI BeGaze version 3.7. Heatmaps are used as

TABLE 1 | Design structure of the loglinear BT pattern model for rankings obtained from $J=5$ images.

Rankings	Paired comparison (PC) patterns										Counts					Model parameters				
	y_{12}	y_{13}	y_{14}	y_{15}	y_{23}	y_{24}	y_{25}	y_{34}	y_{35}	y_{45}	Intercept					x_1	x_2	x_3	x_4	x_5
abcde	1	1	1	1	1	1	1	1	1	1	n_1	1	4	2	0	-2	0	-2	0	-4
bacde	-1	1	1	1	1	1	1	1	1	1	n_2	1	2	4	0	-2	0	-2	0	-4
cabde	-1	-1	1	1	-1	1	1	1	1	1	n_3	1	2	0	4	-2	4	-2	0	-4
...
cedba	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	n_{118}	1	-4	-2	4	0	-2	4	0	2
decba	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	n_{119}	1	-4	-2	0	-4	-2	0	4	2
edcba	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	n_{120}	1	-4	-2	0	-4	-2	0	2	4

Rankings are transformed into paired comparison (PC) patterns; the y 's represent obtained PCs ($y_{jk} = 1$ if $j > k$ and $y_{jk} = -1$ if $k > j$), each possible combination of $J!$ is then counted as observed frequencies in column "counts," and x 's are auxiliary variables used to estimate model parameters indicating how often j was preferred minus how often j was not preferred.

exploratory tools to investigate eye movements (Bojko, 2009) supplementing the BT models.

Data Analytic Strategy

We used Bradley–Terry (BT) models as the basis for recursive partitioning. The BT model is a probability model that can be used to predict the outcome of paired comparisons and to obtain (cardinal) preferences values for all items (images) on a latent scale (Bradley and Terry, 1952). Here, "preference" refers to the judgment of image characteristics (e.g., abstractness) by each participant. Under this model one considers a set of J objects which are presented in pairs. The probability of preferring item j over item k can be described as

$$p_{j>k} = \frac{\pi_j}{\pi_j + \pi_k}, \quad (1)$$

with $\pi_j \geq 0$ and $\sum_{j=1}^J \pi_j$ representing "worth" of the item j , quantifying the position of the item j on a standardized latent scale from 0 to 1. BT models can be fitted as loglinear Bradley–Terry models (LLBT; Sinclair, 1982; Ditttrich et al., 1998). In the basic LLBT, the linear predictor η is given by

$$\eta_{jk} = \ln[m(y_{jk})] = \mu_{jk} + y_{jk}(\lambda_j - \lambda_k), \quad (2)$$

where m denotes the expected frequency of PC decisions, μ_{jk} is a nuisance parameter for the comparison jk which fixes the marginal distribution to n_{jk} and y_{jk} are indicator variables with value 1, if object j is preferred to k and value -1 , if object k is preferred to j . The λ parameters can be transformed into worth parameters by the equation

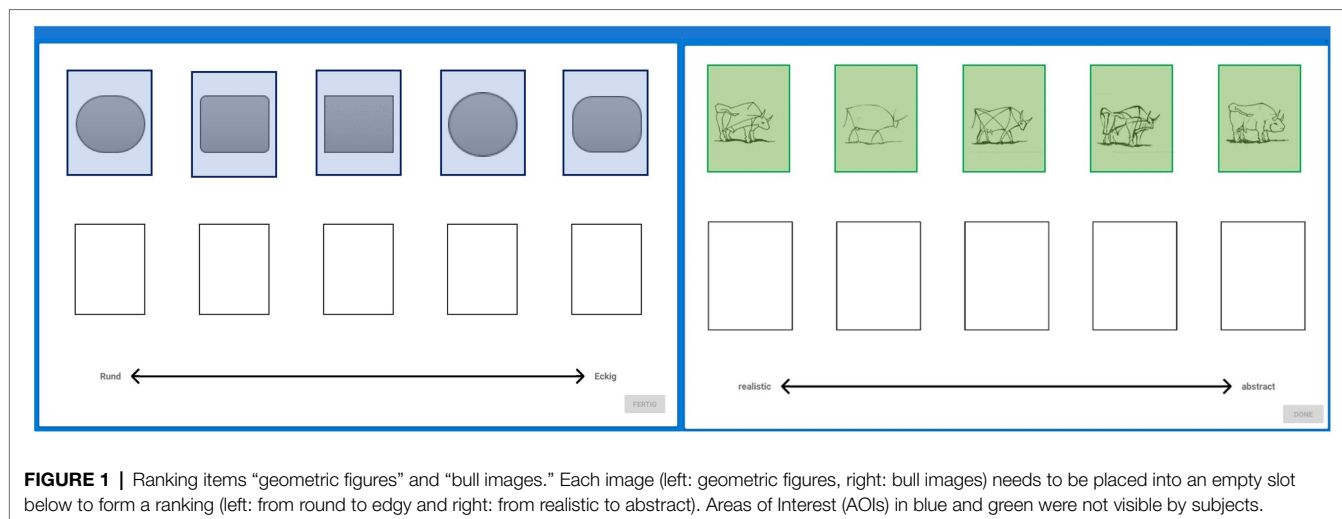
$$\pi_j = \exp(2\lambda_j) / \sum_k \exp(2\lambda_k). \quad (3)$$

As the ranking responses of a subject are considered simultaneously a pattern approach is used. The response pattern is defined as $\mathbf{y} = (y_{12}, y_{13}, \dots, y_{jk}, \dots, y_{J-1,J})$. The expected frequency for a sequence of preferences \mathbf{y} , formulated as a loglinear model, is given as

$$m(\mathbf{y}) = m(y_{12}, \dots, y_{J-1,J}) = np(\mathbf{y}), \quad (4)$$

where n is the total number of respondents and $p(\mathbf{y})$ denotes the probability to observe the response pattern \mathbf{y} .

To gain PC patterns of rankings, rankings are converted into a series of paired comparison decisions (Ditttrich et al., 1998). Note that in the case of forced rankings (i.e., no mid-ranks), ties do not occur by definition. Rankings are transformed into a series of paired comparisons of which intransitive patterns (e.g., $1 > 2$ and $2 > 3$, but $3 > 1$) cannot occur and as such are reduced to $J!$ possible combinations (Ditttrich et al., 2002). Model parameters are estimated using



a log link and a Poisson-distributed error component. **Table 1** shows the design structure of the LLBT model.

To incorporate subject covariates in BT models we used model-based recursive partitioning (MOB; Zeileis et al., 2008) to identify groups of subjects that differ in their preference rankings. The covariate space is recursively divided (partitioned) into sub-groups of subjects with varying image rankings to form a tree-structured division (Strobl et al., 2011). Each terminal node of the tree structure consists of a separate LLBT model with partition-specific model parameters. Wiedermann et al. (2021) extended the MOB BT framework to distinguish between focal independent variables (e.g., expertise status) and covariates used for recursive partitioning. The MOB LLBT model for $g = 1, \dots, G$ subgroups can be written as

$$\log \left[m(y_{jk})_{(g)} \right] = \mu_{(g)} + \lambda_{s(g)} + y_{jkl s(g)} \left(\lambda_{j(g)} + \lambda_{js(g)} - \lambda_{k(g)} - \lambda_{ks(g)} \right), \quad (5)$$

where the intercept $\mu_{(g)}$ and the main effect $\lambda_{s(g)}$ constitute normalizing constants in subgroup g , $y_{jkl s(g)}$ gives the paired comparison decision in group s and partition g (with $y_{jkl s(g)} = 1$ if $j \succ k$ and $y_{jkl s(g)} = -1$ if $k \succ j$), $\lambda_{j(g)}$ and $\lambda_{k(g)}$ denote the partition-specific object parameters for the reference group, and $\lambda_{js(g)}$ and $\lambda_{ks(g)}$ are the partition-specific effects capturing potential group differences (cf. Wiedermann et al., 2021).

Covariates are included to assess the additive impact of subjects' characteristics on the perceived worth of image features. Students in sample I include the following covariates: the time spent on each image set (“Game Time”), gender, age, art grade, and the questions regarding artistic ability and self-perceived art skills. Sample II covariates included age, gender, time spent on each image set, and eye-tracking variables fixation time (time spent fixating image AOIs) and fixation counts (fixations lying inside image AOIs). VL

expertise status (expert vs. novice) served as a focal independent variable.

Statistical analysis and model formulation were conducted with the R-package “prefmod” (Hatzinger and Ditttrich, 2012), partitioning was accomplished with the R-package “partykit” (Hothorn and Zeileis, 2015). To overcome the risk of spurious tree structures a minimum node size of 40 was chosen for Sample I and a minimum of four participants for Sample II to reduce model complexity. To avoid overfitting, a post-pruning strategy based on the Akaike Information Criterion (AIC) was used to prune splits (i.e., bifurcations) that do not improve model fit (Zeileis et al., 2008). Nonparametric bootstrapping (using 1,000 resamples) was used to evaluate the stability of LLBT trees (Philipp et al., 2018). Here, we focused on selection probabilities and average cut-off (splitting) values of the pre-defined covariates. For a stable LLBT tree, selection probabilities of the initially selected covariates are expected to be close to one and average splitting values are expected to be close to the estimates obtained in the initial LLBT tree.

RESULTS

Student Sample I

Table 2 shows the descriptive statistics for self-reported variables and time spent on each image set for sample I. Depending on the image set, different variables had significant impact on the preference rankings.

Table 3 shows the worth parameters for the LLBT tree terminal node in each image set, including significant splitting covariates for sample I. Worth parameters (π) range from 0 to 1, and sum up to 1 for each node. For most image sets, exception being the “salt packages” and the “bull images,” worth parameters decline and form a slope from highest worth to lowest worth according to the intended solution for each image set.

Note that at first glance, certain image sets with worth parameters close to zero would indicate no preference for any of these images. However, this is due to the continuous transformation of the BT model parameters (λ) into a worth parameter (π) on a scale from 0 and 1. For example, for the image set “geometric figures,” each image in the first terminal

node ($n=634$ students) is about 12–20 times more likely to be judged to be more “round” compared to the preceding image in the order “a then b then c then d then e.” Image c, ($\hat{\pi}_c=0.005$) is about 82% more likely to be chosen before image d ($\hat{\pi}_d=0.00041$) from participants in the first terminal node.

Overall, the time spent on each set and the participants’ age had the largest impact on the perceived image features. In general, faster and older student groups tend to form the steepest decline in worth parameters between each image, i.e., image preferences between each image are more clearly separated, indicating no problems in ranking the images according to the intended features. Interestingly, two self-reported visual skills “Interest in visual puzzles” (IP) and “long-term memory” (LM) were important for the judgment of abstraction (i.e., ranking images from realistic to abstract) on item set “dogs” and item set “Mondrian trees.” Here, subgroups with higher scores tended to show steeper decline in worth parameters.

Figure 2 shows the partitioning tree for the dog images. The worth parameter is presented on a log-scale. The student sample is split between fast and slow student groups (about 50%) with one group spending less than 20s on the image set (Game_Time<20) and the other group going above 20s. The gap in perceived abstraction level between dog image b and c is less noticeable for students in node 6 and 7, i.e., slower student groups show similar worth parameters between the two images. However, slower students (45%) with an interest in visual puzzles (IP>1) perceive image c to be less realistic than image b.

TABLE 2 | Descriptive statistics of variables in sample I ($N=987$ students).

Variable	Mean (SD)	
Age	15.35 (2.96)	
S1	3.63 (0.97)	
S2	3.70 (0.89)	
S3	3.33 (0.95)	
S4	3.70 (1.08)	
S5	3.26 (1.16)	
PM	2.57 (0.88)	
SO	3.20 (0.76)	
LM	2.70 (0.8)	
IM	2.05 (0.93)	
IP	2.74 (0.91)	
Art grade	1.96 (0.84)	
Mean time on...		Percentage of correct*
		ranking
Geometric figures	13.28 (5.45)	96%
Dogs	23.01 (10.26)	42%
Bull images	24.33 (12.71)	29%
Mondrian trees	18.16 (9.05)	36%
Salt packages	27.46 (14.49)	04%

S1–S5, self-perceived art skills; PM, photographic memory; SO, spatial orientation; LM, long-term memory; IM, imagination; IP, interest in visual puzzles. *Intended ranking: $a>b>c>d>e$.

TABLE 3 | Worth parameters in each terminal node from sample I.

Sample I—students ($n=987$)							
Image set	Term. node	Worth parameters (π) for each image					Splitting covariates
		a	b	c	d	e	
Geometric figures	$n=634$	0.933	0.061	0.005	4.10E-04	2.00E-05	Age ≤ 15
	$n=259$	0.921	0.069	0.007	9.00E-04	6.30E-05	Age > 15,
	$n=94$	0.593	0.228	0.106	0.053	0.018	Time ≤ 15 s
Dogs	$n=182$	0.415	0.241	0.143	0.120	0.081	Age > 15,
	$n=312$	0.318	0.237	0.184	0.144	0.117	Time > 15 s
	$n=46$	0.280	0.233	0.230	0.158	0.099	Time ≤ 20 s, IP ≤ 2
	$n=447$	0.403	0.223	0.184	0.116	0.074	Time > 20 s, IP > 1
Bull images	$n=76$	0.403	0.226	0.157	0.134	0.080	Time ≤ 12 s
	$n=911$	0.585	0.186	0.091	0.099	0.038	Time > 12 s
Mondrian trees	$n=59$	0.577	0.157	0.135	0.073	0.058	Time < 13 s,
	$n=117$	0.509	0.176	0.182	0.081	0.053	Age ≤ 14
	$n=158$	0.831	0.077	0.074	0.013	0.004	Time < 13 s,
Salt-packages	$n=654$	0.624	0.144	0.136	0.052	0.043	Age > 14, LM ≤ 2
	$n=450$	0.274	0.325	0.144	0.127	0.130	Time < 13 s,
	$n=495$	0.325	0.347	0.113	0.109	0.107	Age > 14, LM > 2

IP, interest in visual puzzles; LM, “I can remember small details in pictures” from 1 (strongly disagree) to 4 (strongly agree).

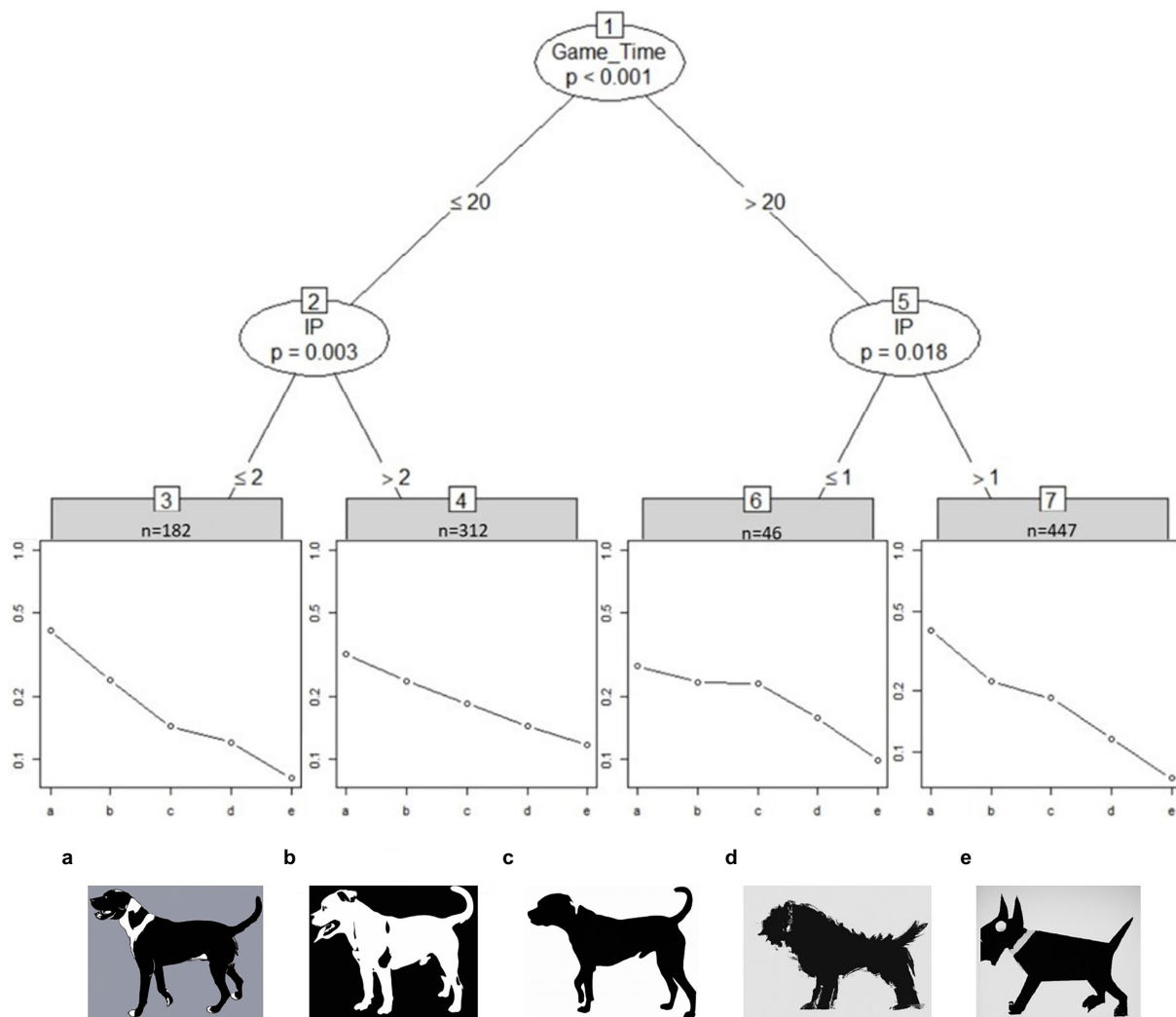


FIGURE 2 | Partitioned paired comparison tree for the ranking task “dogs” in sample I. Game_Time = Time spent on image set in seconds, IP, “Interest in visual puzzles.” Fast students (<20 s) show greater differentiating skill between dog image b and c than slow students (>20 s). Self-reported IP scored greater than 1 increases the perceived differences between dog image b and c in slower student groups (node 7). Placeholder images of dogs due to copyright. Original images can be found at Billmeyer (2017).

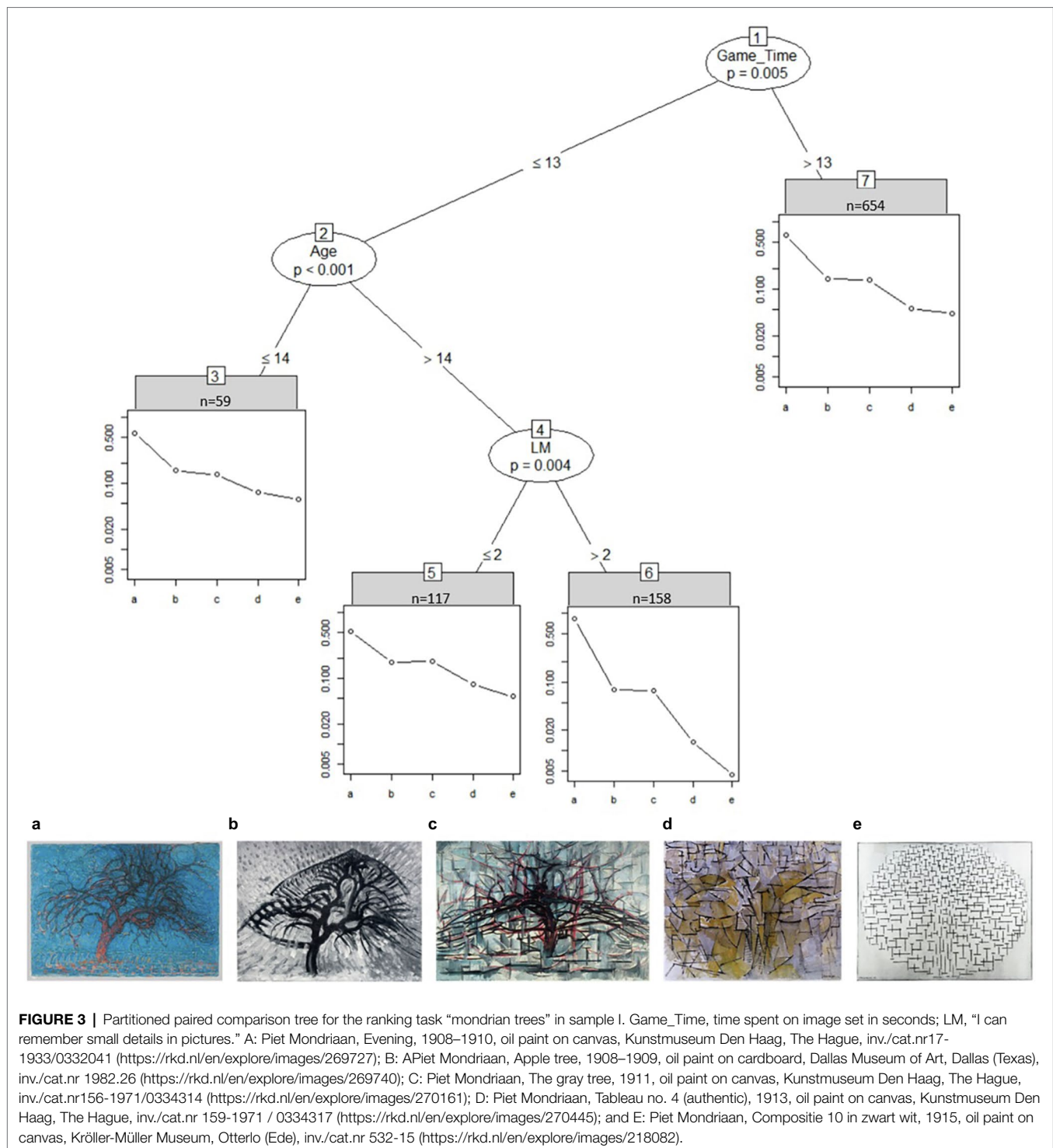
Figure 3 shows how time spent on the task significantly affects the way students in sample I ranked the tree images from realistic (left) to abstract (right). Most students took longer than 13 s to rank the images ($n=654$ in node 7) and ranked images b and c close to each other. Faster students under the age of 15 also ranked the tree images according to their proposed level of abstraction (node 3). Older students with self-reported low long term visual memory skill (LM; disagreeing to the statement “I can remember small details in pictures”) rate image c to be more realistic than image b (node 5). When these students were agreeing or strongly agreeing to that statement instead (node 6) they rated the first image (a) to be nearly 11 times more realistic than the second image (b) and the last image (e) to be about three times more abstract than the fourth image (d).

Figure 4 shows the partitioned tree for the “bull images” set for sample I. Surprisingly, most students (92%) took longer than 12 s and rated image d to be more realistic than image c. The “bull image” set is the only image set with a clear deviation from the intended solution.

Figure 5 shows how the cost of salt packages is clearly split between images a and b vs. c, d, and e. There is also a significant difference in gender: contrary to the actual solution both genders agree b is the most expensive, but males have a smoother drop-off across $a > c > d > e$, whereas females rate a and b as similarly expensive, and c, d, and e as similarly cheap.

Robustness

Stability checks were performed with a bootstrapping procedure, using 1,000 bootstrap samples. **Table 4** shows the probability of splits based on each covariate in sample I and sample II. In sample



I, usually, the time spent on each image set was a common splitting variable, oftentimes splitting the decision tree on each image set except for the “Geometric figures.” Students’ age had significant influence on the stimuli “bull images” and the “Mondrian trees.”

The stability checks indicate that the results from the empirical sample I are comparable: multiple splits on the same decision tree are frequently caused by the time spent on each image set.

The covariates emerging in numerous bootstrap samples exert a more stable impact on the BT model than covariates that emerge only rarely. Questionnaire items S1-S5 on self-reported artistic ability do not seem to trigger splits very often. A few exceptions are noticeable: for the “Mondrian trees” the self-reported ability to imagine (IM) was observed more often to cause a split ($M=0.61$) in comparison to the long-term working memory (LM) variable

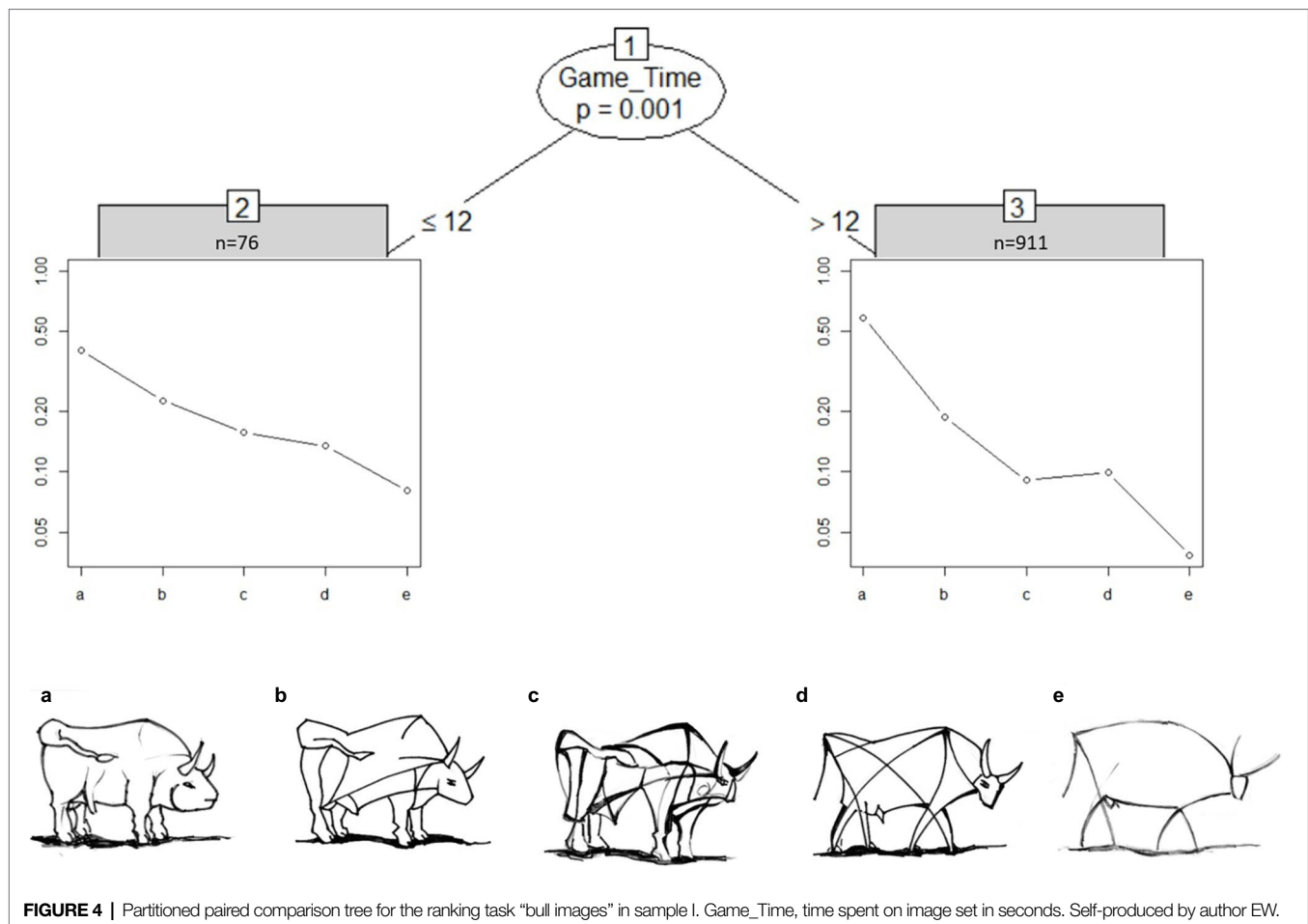


FIGURE 4 | Partitioned paired comparison tree for the ranking task "bull images" in sample I. Game_Time, time spent on image set in seconds. Self-produced by author EW.

($M=0.44$) that is reported in the empirical sample. IM was also nearly equally often used to split the tree of the "Salt packages" stimuli. Additionally, interest in visual puzzles (IP) was also found to split variables on the "bull images" and "Mondrian trees" (>60%), therefore might being underrepresented by the empirical sample. Bootstrapping results for the expert and novices in sample II indicate low splitting probabilities (<15%) for the eye-tracking variables. An exception being the "dogs" image set with fixations on the most realistic image splitting the tree in about 40% of the time. Lastly, the time spent on the dog images was significant in about 50% of the cases.

Figure 6 shows at which values continuous variables split the tree structure as a result of the bootstrapping procedure exemplified for the "bull images" and "Mondrian trees" image set in sample I. For the variable age most splits occurred for students above or below the age of 15 years. The time spent on the task varied for the bull images with a tendency to split at 5 s or between the 10–15 s. Whereas for the "Mondrian trees" splitting peaked around the 7-s mark and then continuously dropped until reaching zero at around 22 s.

Expert-Novice Comparison in Sample II

Worth parameters for the expert and novice comparison are listed in **Table 5**. Generally, experts showed a steeper, linear

decline in worth parameters than novices. Subjects could not be grouped based on the number of fixations and the fixation duration on AOIs. Further, age was the only significant splitting variable on the "bull images" set.

We take a closer look at how this item was perceived by the experts and novices. MOB LLBT results in **Figure 7** indicate that experts above age 28 judge bull image "c" and "d" to be very close in level of abstraction. In contrast, novices above the age of 28 estimate all bulls to have the same distance of abstraction to each other, however this may be due to the small sample size of only three novices in node 3. On the other hand, younger experts show a clear distinction between the most realistic and most abstract bull image, but differentiate only marginally between the three bull images in the middle. Novices below the age of 29 only differentiate strongly between the most realistic bull image to the rest. Generally, older participants differentiate better between the images.

Next, we focus on the distribution of attention for the preference ranking through a fixation heatmap. The mean fixation time spent on the "bull image" set in sample II was $M_{\text{Experts}}=18.37\text{ s}$ ($SD=10.17$), $M_{\text{Novices}}=18.06\text{ s}$ ($SD=8.38$). Repeated ANOVA showed that experts' and novices' fixation times did not significantly differ between each bull [$F(1,47)=0.013$, n.s.]. A comparison of the distribution of fixations on each separate

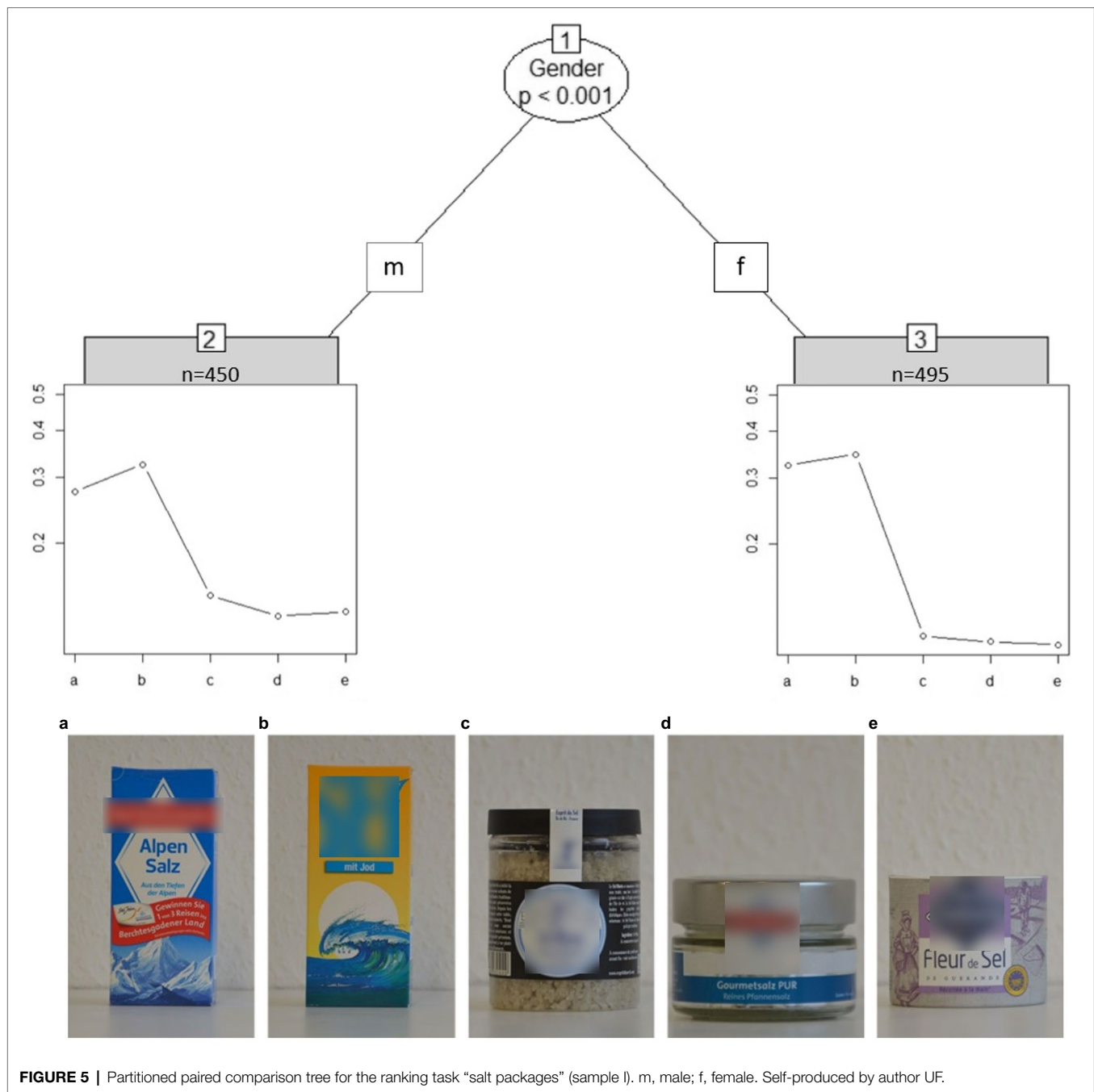


FIGURE 5 | Partitioned paired comparison tree for the ranking task “salt packages” (sample I). m, male; f, female. Self-produced by author UF.

bull image during task completion revealed longer fixation times on bull images b, c and d compared to the most realistic (a) and most abstract (e) bull, $F(4,188)=28.124$, $p<0.001$.

Figure 8 shows a heatmap of mean fixation durations on each bull AOI from start until end of trial, supplementing the model described in **Figure 7**. The most abstract (right) and most realistic (left) bull image attract less attention compared to bulls of similar abstraction level. Fixation times of experts and novices was mainly spent on the bulls associated with a medium level of abstraction (b, c and d). There is a negative correlation between age and fixation time; $r(47)=-0.36$, $p=0.011$, i.e., older participants,

spend less time on images compared to younger participants. Participants below 28 years spend additional fixation time on the most abstract bull image e compared to older groups.

DISCUSSION

This study explored how lay students, lay adults, and visual art experts ranked more or less abstract images by applying a LLBT model to identify potential heterogeneity in visual judgments. Overall, time to complete the ranking task in

TABLE 4 | Selection probabilities of splits for each variable on each image set for bootstrapping procedure on sample I and sample II.

Variable	Probability to split tree				
	Geometric figures	Dogs	Bull images	Mondrian trees	Salt packages
Sample I (n = 987 students)					
Age	0.14	0.49	0.60	0.65	0.45
Gender	0.08	0.42	0.79	0.52	0.92
Game time	0.32	0.98	0.98	0.92	0.82
Art grade	0.16	0.33	0.35	0.39	0.31
S1	0.11	0.25	0.41	0.35	0.30
S2	0.08	0.44	0.52	0.52	0.43
S3	0.28	0.30	0.33	0.30	0.22
S4	0.12	0.27	0.45	0.55	0.42
S5	0.02	0.34	0.29	0.48	0.29
PM	0.29	0.48	0.42	0.44	0.34
SO	0.02	0.40	0.39	0.56	0.45
LM	0.27	0.34	0.42	0.44	0.33
IM	0.11	0.54	0.53	0.61	0.62
IP	0.18	0.82	0.62	0.66	0.45
Sample II (n = 49 VL-experts and novices)					
Age	0.00	0.20	0.40	0.17	–
Gender	0.00	0.01	0.00	0.00	–
Game time	0.00	0.48	0.03	0.06	–
Fix. duration a	0.00	0.19	0.07	0.05	–
Fix. duration b	0.00	0.03	0.00	0.00	–
Fix. duration c	0.00	0.04	0.00	0.05	–
Fix. duration d	0.00	0.02	0.01	0.15	–
Fix. duration e	0.00	0.06	0.01	0.00	–
Fix. count a	0.00	0.39	0.01	0.01	–
Fix. count b	0.00	0.28	0.01	0.00	–
Fix. count c	0.00	0.08	0.00	0.03	–
Fix. count d	0.00	0.09	0.05	0.01	–
Fix. count e	0.00	0.05	0.02	0.00	–

Probabilities of splits > 0.60 are marked in bold. S1–S5, self-perceived art skills; PM, photographic memory; SO, spatial orientation; LM, long-term memory; IM, imagination; IP, interest in visual puzzles; a = most realistic image to e = most abstract image.

combination with self-reported skills have significant influence on model parameters. In general, the longer students took to rank the images, the closer each image was ranked to the previous one, i.e., the difference in the ranked preferences between the images decreases. Students who spent more time on the task may have had difficulties ranking the images the intended way. Additionally, visual skills affected the ease to differentiate between images. Interestingly, the students' art grade did not affect the ability to rank the presented images with respect to visual abstraction. There was also no apparent classroom group effect.

The slim packaging of the “salt packages” seems to determine the perceived difference in cost. In contrast to other images, the knowledge of goods and prices is very different to the evaluation of image abstraction and is well reflected by the preference scale: the divergence between small and round vs. slim and tall salt packaging can be clearly seen in the steep drop of estimated worth parameters after image “b.” It could be hypothesized that male and female students might have different access to merchandise, which could explain the slight difference in cost perception by gender.

Furthermore, ranking abstract images such as the “bull images” revealed how similar abstraction levels of image pairs are reflected by similar worth parameters. The majority of students ranked bull image d as more realistic even though it contains less features than c. Apparently line thickness influences the perception of abstraction level for the majority of students. Also, the bull's eye is drawn slightly more realistically in bull d in comparison to bull c, which may have influenced the ranking. Are these differences in perceived judgment of images outside the intended ranking an indication for less skilled student groups? This cannot be derived solely from the ranked preferences. Comparing this result to the sample II, revealed how VL experts above the age of 28 judged both bull images c and d to be nearly identical in abstraction level. Exploring the fixation distribution of VL experts' and novices' eye movements, exemplified by heatmaps, showed how images of similar abstraction level (with similar worth parameters) evoke longer fixation durations.

Students with high self-reported interest in visual puzzle solving were able to distinguish abstract images more clearly. The self-reported ability to remember small details in pictures (“working memory”) also contributed to students' ability to rank the level of abstraction of the images, indicated by greater systematic difference (i.e., exhibiting a steeper slope across the five images) in worth parameters between each image pair. Stability checks suggest that MOB LLBT models can sufficiently detect heterogeneity of visual judgments in a large sample of students. The time students took to rank the images was a significant splitting covariate for almost all image sets. The interest in visual puzzles was the most relevant self-reported ability for ranking abstract images. Furthermore, age, for example, was a less prevalent splitting variable for the “dogs” image set but not for the “bull images” and “Mondrian trees.” This might be caused by the difference between abstraction due to signal character (dogs as information) vs. an aesthetic expression (trees and bulls as illustrations of experiences).

As seen in the results of the expert and novice comparison in sample II, VL experts were able to determine nuanced abstraction levels between images, as reflected in the similar worth parameters between image pairs. Smaller differences between certain image pairs do not necessarily reflect poorly on the ability to differentiate abstract images, but may indicate subtle image variations perceived by experts. Thus, especially when dealing with images of artwork, an interpretation by art experts and teachers is advisable.

LIMITATIONS

A few limitations of the present study should be mentioned. Firstly, as an exploratory study by design, generalizability of empirical results is limited. Only a reduced number of item sets were presented. Causal effects of covariates over different stimuli would require an experimental design that systematically varies visual stimuli and should be tested at the end of a longer series of experimental studies. Even though the intended ranking for abstract images was moderately low (between 29% and 42%), the worth parameters did not reflect the presence of outlying responses between student groups, i.e., there was no large systematic difference in ranking order among students.

Different sets of stimuli, e.g., computer generated art that controls for salience (Furnham and Rao, 2002; Shakeri et al., 2017) with a focus on a single dimensions of visual abstraction, such as composition or color (Markovic, 2010) could lead to higher variability in perceived judgment.

In comparison to other image ranking tasks (e.g., Strobl et al., 2011), an intended ordering of items was agreed upon. In the case of latent image characteristics multiple orderings may be acceptable and should be elaborated upon further (such as in the case of the “bull images”). However, a ranking assignment with heterogeneous preference patterns might indicate ambiguities with selected items. For educational assignments a clear preference ranking, with uniformly distributed worth parameters might be more desirable.

In sample II only age was found as a significant splitting variable, which might be due to low statistical power. Age of

participants might also be confounded with expertise as older persons tend to have more expertise. Finally, the number of datapoints increase dramatically with the number of items for MOB LLBT models. With $5! = 120$ possible PC patterns and $n = 987$ participants, the resulting input dataset consists of 118,440 observations, owing to the separate design matrices for each subject. Researchers might consider limiting the number of items during study design to reduce the design complexity.

CONCLUSION

As an empirically derived observation our results suggest the following: less time spent on the visual judgments was associated with the ability to better discriminate between images of varying levels of abstraction. Abilities related to visual arts (imagination

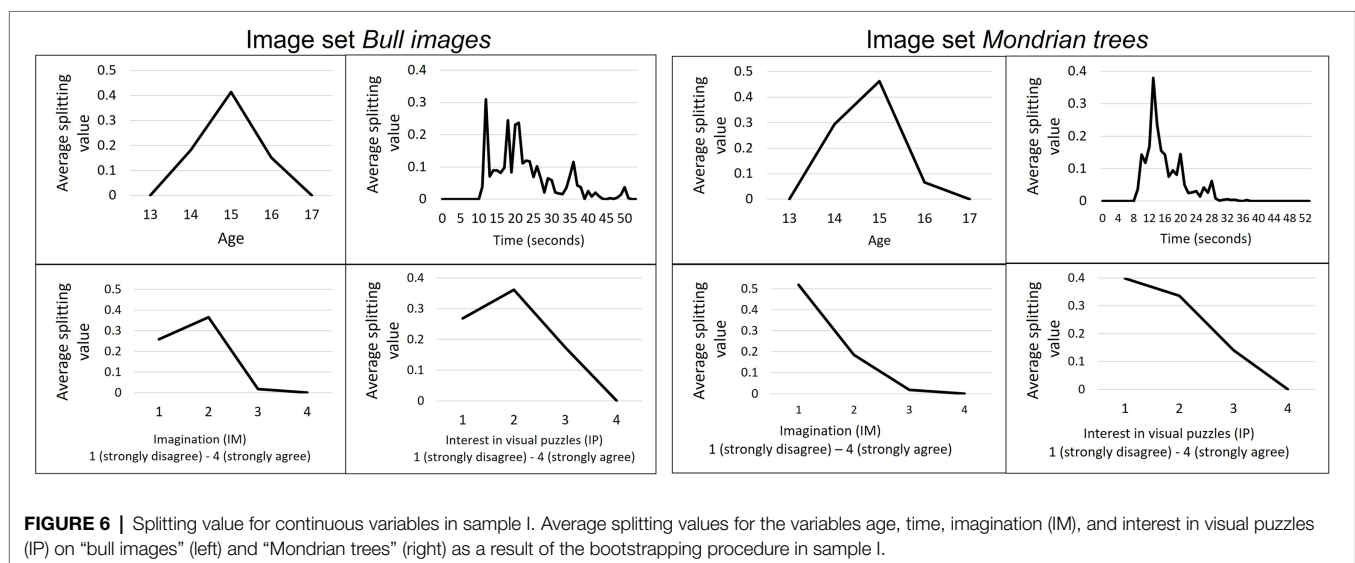
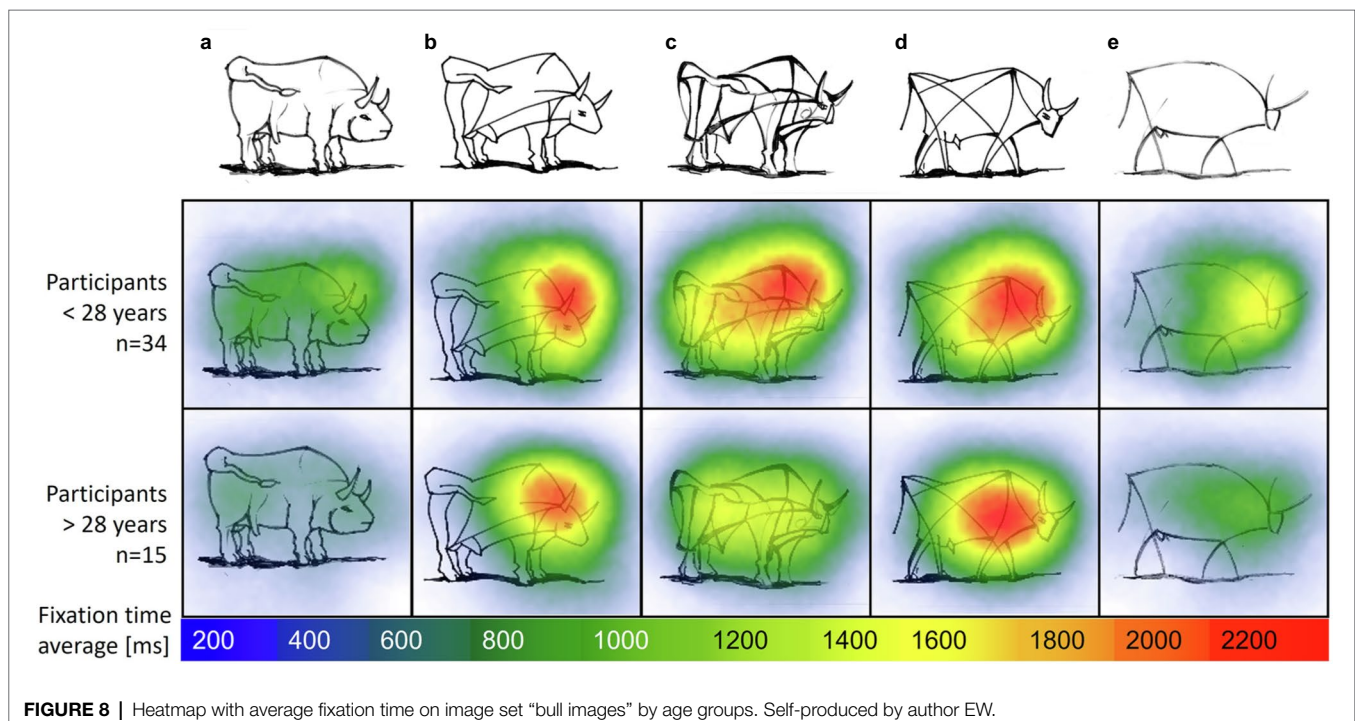
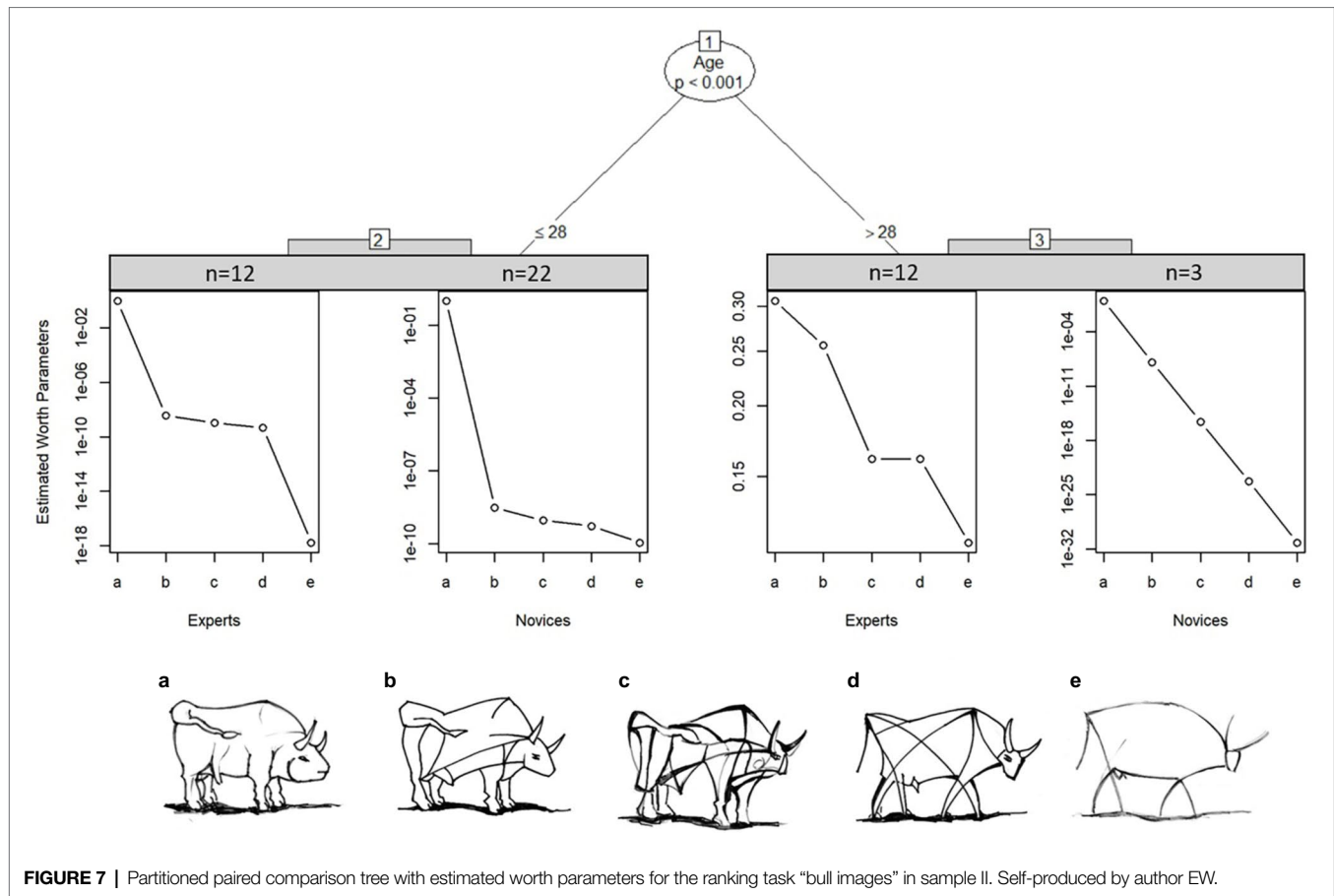


TABLE 5 | Worth parameters in each terminal node from sample II.

Sample II—VL experts and novices ($n = 49$)							
Image set	Term. node	Worth parameters (π) for each image (95% CI)					Splitting covariates
		a	b	c	d	e	
Geo-metric figures	$n = 24$ Exp	0.999 (0.99–0.99)	1.65e-09 (6.4e-10–4.2e-09)	4.65e-18 (9.5e-19–2.2e-17)	1.31e-26 (1.21e-27–1.41e-25)	2.17e-35 (9.5e-38– 4.9e-33)	–
	$n = 25$ Nov	0.608 (0.48–0.85)	0.248 (0.12–0.40)	0.089 (0.04–0.19)	0.043 (0.01–0.08)	0.012 (0.004–0.06)	
Dogs	$n = 24$ Exp	0.325 (0.31–0.34)	0.255 (0.25–0.26)	0.191 (0.19–0.19)	0.135 (0.13–0.13)	0.095 (0.08–0.12)	–
	$n = 25$ Nov	0.478 (0.42–0.52)	0.197 (0.19–0.19)	0.168 (0.17–0.16)	0.100 (0.09–0.10)	0.057 (0.03–0.11)	
Bull images	$n = 12$ Exp	0.999 (0.99–0.99)	3.57e-09 (3.5e-09–3.6e-09)	1.05E-09 (1.1E-09–1.1E-09)	4.55e-10 (4.3e-10–4.7e-10)	1.62e-18 (1.2e-18–2.3e-18)	Age ≤ 28
	$n = 22$ Nov	0.999 (0.99–0.99)	3.01e-09 (2.6e-09–3.4e-09)	9.28E-10 (7.5e-10–1.1e-09)	5.25e-10 (3.9e-10–7.0e-10)	1.1e-10 (4.4e-11–2.7e-10)	
	$n = 12$ Exp	0.307 (0.29–0.32)	0.256 (0.25–0.26)	0.161 (0.16–0.16)	0.161 (0.16–0.16)	0.114 (0.09–0.15)	
	$n = 3$ Nov	0.999 (0.99–0.99)	1.22e-08 (2.6e-07–3.4e-09)	2.54e-16 (7.5e-15–2.83-16)	5.2e-24 (7.0e-20–3.9e-28)	6.45e-32 (4.4e-30–2.7e-34)	Age > 28
	$n = 24$ Exp	0.748 (0.70–0.78)	0.141 (0.12–0.15)	0.085 (0.07–0.09)	0.021 (0.01–0.03)	0.006 (0.002–0.011)	
	$n = 25$ Nov	0.999 (0.99–0.99)	2.21E-08 (1.9e-08–2.5e-08)	1.10E-08 (8.9e-09–1.4e-08)	2.58E-09 (1.4e-09–4.8e-09)	9.61E-10 (1.7e-10–5.2e-09)	
Mondrian trees							



and interest in visual puzzles) seem to support this discriminative ability demonstrated by our participants.

In contrast to measurements of visual judgment with visual analog scales (e.g., the AAA instrument by Chatterjee et al. (2010)), ranking tasks lets participants compare multiple images at once. BT trees then can be used in various educational settings, e.g., art assignments where exact iconicity between two images is unknown. Judging images of more varying complexity (see García et al., 1994 for an early attempt to measure icon complexity) could be a next step in the construction of future test batteries on VL.

The presented modelling approach allows one to quantify the distance between images on a standardized latent scale. Here, BT models do not rely on the assumption of equidistant response categories. The latent metric scale is derived from ordinal (ranking) data to capture the perceived between-group differences of visual judgment. The perceived distance between each image (e.g., level of abstraction) can then be used to identify closely related and, therefore, hard-to-differentiate objects. Such objects could subsequently be discussed and analyzed in art class.

AUTHOR'S NOTE

The results of this study were used by author MT to fulfil some of the requirements for the doctoral degree program at the University of Regensburg.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

REFERENCES

- Andrews, K., Zimoch, M., Reichert, M., Tallon, M., Frick, U., and Pryss, R. (2018). A smart mobile assessment tool for collecting data in large-scale educational studies. *Procedia Comput. Sci.* 134, 67–74. doi: 10.1016/j.procs.2018.07.145
- Billmeyer, F. (2017). Bilder – mehr oder weniger ähnlich – Hunde. Bilderlernen. Available at: <https://www.bilderlernen.at/2017/11/18/bilder-mehr-oder-weniger-aehnlich/> (Accessed November 18, 2017).
- Bojko, A. (2009). "Informative or misleading? Heatmaps deconstructed," in *Lecture Notes in Computer Science: Vol. 5610. Human-Computer Interaction. New Trends. HCI 2009. Lecture Notes in Computer Science*. ed. J. A. Jacko (Berlin, Heidelberg: Springer), 30–39.
- Bradley, R. A., and Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika* 39, 324–345. doi: 10.2307/2334029
- Brams, S., Ziv, G., Levin, O., Spitz, J., Wagemans, J., Williams, A. M., et al. (2019). The relationship between gaze behavior, expertise, and performance: a systematic review. *Psychol. Bull.* 145, 980–1027. doi: 10.1037/bul0000207
- Bromberger, E. (2021). The potential of eye tracking for visual literacy research. *Journal of Visual Literacy* 40, 34–50. doi: 10.1080/1051144X.2021.1902040
- Cattelan, M. (2012). Models for paired comparison data: a review with emphasis on dependent data. *Stat. Sci.* 27, 412–433. doi: 10.1214/12-sts396
- Cattelan, M., Varin, C., and Firth, D. (2013). Dynamic Bradley-Terry modelling of sports tournaments. *J. R. Stat. Soc. Ser. C. Appl. Stat.* 62, 135–150. doi: 10.1111/j.1467-9876.2012.01046.x

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Research of the Leibniz Institute for Research and Information in Education, Frankfurt am Main (DIPE, 01JK1606A). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

MT, UF, and KR designed the study. MT and EW selected and prepared the stimuli. MT conducted the field work and eye-tracking experiments, performed the statistical analysis with input on data interpretation by WW, UF, and MG, and prepared the manuscript. All authors reviewed the article and approved the submitted version.

FUNDING

This work was supported by the German Federal Ministry of Education and Research (BMBF) under grant number 01JK1606A.

ACKNOWLEDGMENTS

We would like to thank all students, VL experts, and novices who participated in this study.

- Chakraborty, H., and Sen, P. K. (2016). Resampling method to estimate intra-cluster correlation for clustered binary data. *Commun. Stat. Theory Methods* 45, 2368–2377. doi: 10.1080/03610926.2013.870202
- Chamorro-Premuzic, T., and Furnham, A. (2004). Art judgment: a measure related to both personality and intelligence? *Imagin. Cogn. Pers.* 24, 3–24. doi: 10.2190/U4LW-TH9X-80M3-NJ54
- Chatterjee, A., Widick, P., Sternschein, R., Smith, W. B., and Bromberger, B. (2010). The assessment of art attributes. *Empir. Stud. Arts* 28, 207–222. doi: 10.2190/EM.28.2.f
- Child, I. L. (1965). Personality correlates of esthetic judgment in college students. *J. Pers.* 33, 476–511. doi: 10.1111/j.1467-6494.1965.tb01399.x
- Choi, B. C. K., and Pak, A. W. P. (2005). A catalog of biases in questionnaires. *Prev. Chronic Dis.* 2:A13.
- Dittrich, R., Hatzinger, R., and Katzenbeisser, W. (1998). Modelling the effect of subject-specific covariates in paired comparison studies with an application to university rankings. *J. R. Stat. Soc. Ser. C. Appl. Stat.* 47, 511–525. doi: 10.1111/1467-9876.00125
- Dittrich, R., Hatzinger, R., and Katzenbeisser, W. (2002). Modelling dependencies in paired comparison data. *Comput. Stat. Data Anal.* 40, 39–57. doi: 10.1016/S0167-9473(01)00106-2
- Dittrich, R., Katzenbeisser, W., and Reisinger, H. (2000). The analysis of rank ordered preference data based on Bradley-Terry type models. *OR Spectr.* 22, 117–134. doi: 10.1007/s002910050008
- Engbert, R., Rothkegel, L. O. M., Backhaus, D., and Trukenbrod, H. A. (2016). Evaluation of velocity-based saccade detection in the SMI-ETG 2W system. Technical Report.
- Francis, B., Dittrich, R., and Hatzinger, R. (2010). Modeling heterogeneity in ranked responses by nonparametric maximum likelihood: how do Europeans

- get their scientific knowledge? *Ann. Appl. Stat.* 4, 2181–2202. doi: 10.1214/10-AOAS366
- Frick, U., Rakoczy, K., Tallon, M., Weiß, S., and Wagner, E. (2020). “Ich sehe was, was Du nicht siehst! Erste Bausteine zur Messung von Bildkompetenz bei Schüler*innen der 9. und 10. Jahrgangsstufe [I can see what you cannot see! First building blocks for measuring visual literacy in 9th and 10th grade],” in *Kulturelle Bildung: Theoretische Perspektiven, methodologische Herausforderungen und empirische Befunde*. eds. S. Timm, J. Costa, C. Kühn and A. Scheunpflug (Münster: Waxmann), 379–399.
- Furnham, A., and Boo, H. C. (2011). A literature review of the anchoring effect. *J. Socio-Econ.* 40, 35–42. doi: 10.1016/j.socec.2010.10.008
- Furnham, A., and Rao, S. (2002). Personality and the aesthetics of composition: a study of Mondrian and Hirst. *N. Am. J. Psychol.* 4, 233–242.
- García, M., Badre, A. N., and Stasko, J. T. (1994). Development and validation of icons varying in their abstractness. *Interact. Comput.* 6, 191–211. doi: 10.1016/0953-5438(94)90024-8
- Gortais, B. (2003). Abstraction and art. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 1241–1249. doi: 10.1098/rstb.2003.1309
- Hatzinger, R., and Dittrich, R. (2012). prefmod: an R Package for Modeling Preferences Based on Paired Comparisons, Rankings, or Ratings. *J. Stat. Softw.* 48, 1–31. doi: 10.18637/jss.v048.i10
- Hayn-Leichsenring, G. U., Kenett, Y. N., Schulz, K., and Chatterjee, A. (2020). Abstract art paintings, global image properties, and verbal descriptions: an empirical and computational investigation. *Acta Psychol. (Amst)* 202:102936. doi: 10.1016/j.actpsy.2019.102936
- Hothorn, T., and Zeileis, A. (2015). partykit: a modular toolkit for recursive partytioning in R. *J. Mach. Learn. Res.* 16, 3905–3909.
- Jacobsen, T. (2004). Individual and group modelling of aesthetic judgment strategies. *Br. J. Psychol.* 1953, 95, 41–56. doi: 10.1348/000712604322779451
- Jarodzka, H., Gruber, H., and Holmqvist, K. (2017). Eye tracking in Educational Science: Theoretical frameworks and research agendas. *J. Eye Mov. Res.* 10, 1–18. doi: 10.16910/jemr.10.1.3
- MacTaggart, J. (2021). Animals in Art-Pablo Picasso. Available at: https://www.artfactory.com/art_appreciation/animals_in_art/pablo_picasso.htm (Accessed October 06, 2021).
- Markovic, S. (2010). Perceptual, semantic and affective dimensions of the experience of representational and abstract paintings. *J. Vis.* 10:1230. doi: 10.1167/10.7.1230
- McCormack, J., Cruz Gambardella, C., and Lomas, A. (2021). “The enigma of complexity,” in *Lecture Notes in Computer Science: Vol. 12693. Artificial Intelligence in Music, Sound, Art and Design. Vol. 12693*. eds. J. Romero, T. Martins and N. Rodríguez-Fernández (Cham: Springer International Publishing), 203–217.
- McManus, I. C., Cheema, B., and Stoker, J. (1993). The aesthetics of composition: a study of Mondrian. *Empir. Stud. Arts* 11, 83–94. doi: 10.2190/HXR4-VU9A-P5D9-BPQQ
- Nodine, C. F., Locher, P. J., and Krupinski, E. A. (1993). The role of formal art training on perception and aesthetic judgment of art compositions. *Leonardo* 26, 219–227. doi: 10.2307/1575815
- Philipp, M., Rusch, T., Hornik, K., and Strobl, C. (2018). Measuring the stability of results from supervised statistical learning. *J. Comput. Graph. Stat.* 27, 685–700. doi: 10.1080/10618600.2018.1473779
- Schoch, K., and Ostermann, T. (2020). Giving the art greater weight in art psychology: Rizba, a questionnaire for formal picture analysis. *Creativity. Theo. – Res. – Appl.* 7, 373–410. doi: 10.2478/ctra-2020-0019
- Shakeri, H., Nixon, M., and DiPaola, S. (2017). Saliency-based artistic abstraction with deep learning and regression trees. *J. Imaging Sci. Technol.* 61, 60402-1–60402-9. doi: 10.2352/J.ImagingSci.Technol.2017.61.6.060402
- Sinclair, C. D. (1982). “GLIM for preference,” in *Lecture Notes in Statistics. GLIM 82: Proceedings of the International Conference on Generalised Linear Models. Vol. 14*. eds. D. Brillinger, S. Fienberg, J. Gani, J. Hartigan, K. Krickeberg and R. Gilchrist (New York: Springer), 164–178.
- Specht, S. M. (2007). Successive contrast effects for judgments of abstraction in artwork following minimal pre-exposure. *Empir. Stud. Arts* 25, 63–70. doi: 10.2190/W717-88W2-2233-12H3
- Streiner, D. L., and Norman, G. R. (2008). “Biases in responding,” in *Health Measurement Scales: A Practical Guide to Their Development and Use. 4th Edn.* eds. D. L. Streiner and G. R. Norman (Oxford: Oxford University Press), 103–134.
- Strobl, C., Wickelmaier, F., and Zeileis, A. (2011). Accounting for individual differences in Bradley-Terry models by means of recursive partitioning. *J. Educ. Behav. Stat.* 36, 135–153. doi: 10.3102/1076998609359791
- Vansteenkiste, P., Cardon, G., Philippaerts, R., and Lenoir, M. (2015). Measuring dwell time percentage from head-mounted eye-tracking data—comparison of a frame-by-frame and a fixation-by-fixation analysis. *Ergonomics* 58, 712–721. doi: 10.1080/00140139.2014.990524
- Viola, I., Chen, M., and Isenberg, T. (2020). “Visual abstraction,” in *Foundations of Data Visualization*. eds. M. Chen, H. Hauser, P. Rheingans and G. Scheuermann (Cham: Springer International Publishing), 15–37.
- Wagner, E., and Schönaue, D. (eds.) (2016). *Common European Framework of Reference for Visual Literacy-Prototype*. Münster: Waxmann.
- Wiedermann, W., Frick, U., and Merkle, E. C. (2021). Detecting heterogeneity of intervention effects in comparative judgments. *Prev. Sci.* doi: 10.1007/s11121-021-01212-z
- Wiedermann, W., Niggli, J., and Frick, U. (2014). The lemming-effect: harm perception of psychotropic substances among music festival visitors. *Health Risk Soc.* 16, 323–338. doi: 10.1080/13698575.2014.930817
- Witkin, R. W. (1983). The psychology of abstraction and the visual arts. *Leonardo* 16:200. doi: 10.2307/1574914
- Zeileis, A., Hothorn, T., and Hornik, K. (2008). Model-based recursive partitioning. *J. Comput. Graph. Stat.* 17, 492–514. doi: 10.1198/106186008X319331

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Tallon, Greenlee, Wagner, Rakoczy, Wiedermann and Frick. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



OPEN ACCESS

EDITED BY

Alyssa A. Brewer,
University of California, Irvine,
United States

REVIEWED BY

M. Pilar Aivar,
Autonomous University of Madrid,
Spain
Robert Leslie Whitwell,
Western University, Canada

*CORRESPONDENCE

Ryan W. Langridge
langrirw@myumanitoba.ca

SPECIALTY SECTION

This article was submitted to
Perception Science,
a section of the journal
Frontiers in Psychology

RECEIVED 17 April 2022

ACCEPTED 01 August 2022

PUBLISHED 02 September 2022

CITATION

Langridge RW and Marotta JJ (2022)
Use of remote data collection
methodology to test for an illusory
effect on visually guided cursor
movements.
Front. Psychol. 13:922381.
doi: 10.3389/fpsyg.2022.922381

COPYRIGHT

© 2022 Langridge and Marotta. This is
an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in
other forums is permitted, provided
the original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Use of remote data collection methodology to test for an illusory effect on visually guided cursor movements

Ryan W. Langridge* and Jonathan J. Marotta

Perception and Action Lab, Department of Psychology, University of Manitoba, Winnipeg, MB, Canada

Investigating the influence of perception on the control of visually guided action typically involves controlled experimentation within the laboratory setting. When appropriate, however, behavioral research of this nature may benefit from the use of methods that allow for remote data collection outside of the lab. This study tested the feasibility of using remote data collection methods to explore the influence of perceived target size on visually guided cursor movements using the Ebbinghaus illusion. Participants completed the experiment remotely, using the trackpad of their personal laptop computers. The task required participants to click on a single circular target presented at either the left or right side of their screen as quickly and accurately as possible (Experiment 1), or to emphasize speed (Experiment 2) or accuracy (Experiment 3). On each trial the target was either surrounded by small or large context circles, or no context circles. Participants' judgments of the targets' perceived size were influenced by the illusion, however, the illusion failed to produce differences in click-point accuracy or movement time. Interestingly, the illusion appeared to affect participants' movement of the cursor toward the target; more directional changes were made when clicking the Perceived Large version of the illusion compared to the Perceived Small version. These results suggest the planning of the cursor movement may have been influenced by the illusion, while later stages of the movement were not, and cursor movements directed toward targets perceived as smaller required less correction compared to targets perceived as larger.

KEYWORDS

Ebbinghaus illusion, Titchener circles, cursor control, perception, action

Introduction

The Ebbinghaus illusion, also referred to as the Titchener circles illusion, is a well-known size-contrast illusion in which the perceived size of a central target circle is made to appear smaller or larger than its true size when surrounded by a ring of larger or smaller context circles, respectively. The strength of the illusion can be manipulated by

altering the size and distance of the context circles relative to the target circle; smaller distances between the target circle and the surrounding annulus increase the perceived size of the target circle, while larger distances decrease its perceived size (Massaro and Anderson, 1971; Roberts et al., 2005; Knol et al., 2015). Visual illusions such as the Ebbinghaus illusion provide an opportunity to explore the degree of separation between a visual system dedicated specifically to the processing of a stimulus' perceptual properties, and a visual system dedicated specifically to the execution of visually guided action toward that stimulus. A functional separation of these two visual systems, as proposed by Goodale and Milner (1992) and Milner and Goodale (2006) suggests that a size-contrast illusion such as the Ebbinghaus illusion should primarily influence one's perceptual judgments of a stimulus' size processed within the ventral stream, while the visually guided action toward that stimulus, guided by computations performed by the dorsal stream, should be largely unaffected by the illusory context.

Traditionally, research of this nature involves measuring participants' grip aperture when reaching to grasp a circular disk embedded within the illusion and drawing a comparison to participants' judgments of the disk's perceived size in response to the illusory context. A number of studies have provided evidence supporting the theory that visually guided grasping actions are immune to the influence of the Ebbinghaus illusion by describing relatively stable, non-changing grip apertures in comparison to perceptual size-judgments that vary as a function of the illusory context (Aglioti et al., 1995; Haffenden and Goodale, 1998; Marotta et al., 1998; Danckert et al., 2002). Yet others have provided contradictory evidence suggesting that both perceptual judgements and visually guided action are influenced to some degree by visual illusions (Pavani et al., 1999; Franz et al., 2000; Franz and Gegenfurtner, 2008). Some of these contradictions may result from variations in the presentation of the illusion across investigations (e.g., differing size and presentation of the targets, variable number of surrounding context circles, etc.), as well as variation in the methods used to measure the perceptual influence of the illusion itself: for instance, having the participants adjust the size of a comparison stimulus to match the size of the target stimulus (e.g., Aglioti et al., 1995; Haffenden and Goodale, 1998) vs. using the distance between the index finger and thumb during perceptual estimation of the stimulus size (e.g., Haffenden et al., 2001). Incongruencies in study design such as these may explain why some studies have demonstrated an effect of the Ebbinghaus illusion on visually guided action, while others have not.

Some who argue for an "illusion immunity" of visually guided action contend that the apparent effects of the illusion on grip aperture may be the result of an obstacle avoidance mechanism, suggesting any observed changes in grip aperture are caused by the proximity of the context circles to the target circle, rather than in response to a perceived change in target circle size. Certain studies have found evidence for this

hypothesis (Haffenden and Goodale, 2000; Haffenden et al., 2001; De Grave et al., 2005; Gilster et al., 2006). However, there are others who have provided evidence suggesting the positioning of the context circles is not a sufficient explanation for the observed changes in grip aperture, and therefore these changes must arise in response to a change in the perceived size of the target circle (Franz et al., 2003; Franz and Gegenfurtner, 2008; Kopiske et al., 2016).

The extent to which the Ebbinghaus illusion influences the precision and timing of other visually guided actions such as pointing or tapping has also been explored (van Donkelaar, 1999; Handlovsky et al., 2004; Alphonsa et al., 2016; Knol et al., 2017). Instead of requiring an appropriately scaled grasp matched to the boundaries of a circular target object, these actions require performing an accurate movement toward the target's center, and the influence of the illusion is typically investigated by testing for the presence of a speed-accuracy trade-off (e.g., Fitts' Law; Fitts, 1954), when acting on targets perceived to be smaller or larger than their veridical size. However, as with investigations of grasping the Ebbinghaus illusion, the results of studies involving this type of action are also often contradictory. For example, van Donkelaar (1999) originally demonstrated slower movement times toward targets perceived to be smaller, however, attempts to replicate these results failed to demonstrate an influence of the illusion on movement time (Fischer, 2001). A critical difference between the early van Donkelaar (1999) and later Fischer (2001) studies involved visibility of the hand: participants in van Donkelaar's original study did not have visual feedback of their hand, while those in Fischer's later study were able to see their hand while performing the task. In a similar study, Alphonsa et al. (2016) adapted a typical Fitts tapping task to include qualities of both the Ebbinghaus and the Muller-Lyer illusions. Despite the physical size and distance of the two target stimuli being identical, participants were more accurate when tapping targets in the "illusory easy" condition, in which the combination of illusions increased the perceived size of the target. However, the effect of the illusion on participants' tapping accuracy was only observed during "discrete" tapping, where visual feedback of the target was removed. Together, these results suggest that the degree to which the illusion influences movement time may be related to the amount of visual information available. In instances where an influence of the Ebbinghaus illusion is observed, the evidence appears to suggest increased movement time for actions directed toward stimuli perceived to be smaller than their veridical size.

More than ever before, the research community is now looking for ways in which to adapt classic in-person data collection procedures into formats that allow participant data to be collected remotely. While research utilizing the use of online questionnaires and other forms of self-report measures has long benefited from remote data collection, behavioral research requiring the increased experimental

control provided by the laboratory setting, as well as the use of expensive, often cumbersome equipment, has necessitated these types of experiments to be conducted in the lab, under experimenter supervision. The aim of this study was to test the implementation of remote data collection methods to investigate behaviors such as visually guided movement. In particular, this study measured the extent to which perceptual information influences participants' cursor movements when using a laptop trackpad to perform a point-and-click task, an increasingly common behavior for many people. Despite being far removed from the type of action humans' visual system evolved to facilitate, this type of task provides an interesting context in which to measure the visual perception-action relationship. Onscreen cursor movement requires a transformation from egocentrically defined finger movements—typically controlled via the vision-for-action dorsal stream—into the on-screen environment, where allocentric references are critical for the guidance of the cursor to the desired location and are therefore likely influenced to some degree by the vision-for-perception ventral stream (Goodale and Milner, 1992; Milner and Goodale, 2006). Cursor movements have also been shown to adhere to Fitts' Law (Sutter et al., 2011), suggesting there is a speed-accuracy trade off present following this type of visuomotor transformation as well. While the processes involved in reaching toward and grasping a physical object are inherently different from those which serve interaction with 2-D stimuli (Freud et al., 2018; Ozana et al., 2020), the use of a trackpad to control an onscreen cursor has become an increasingly prevalent behavior for many individuals. As such, exploring the ways in which the visual system performs these transformations is an interesting direction of study.

If the transformation from finger movement into cursor movement requires some degree of perceptual control, we may expect one's cursor movements to be influenced by their perception of the onscreen stimuli being clicked on. Using the Ebbinghaus illusion to influence participants' perceptions of target size, we tested the degree to which participants' cursor movements toward targets embedded within the Ebbinghaus illusion were influenced by the illusory context. In doing so, we incorporated features of a classic behavioral investigation into an experiment that could be conducted remotely, using the participants' own laptop device. The results of three experiments are detailed here, along with a discussion of the potential strengths and limitations present when conducting this type of experiment remotely.

Methods

Participants

All participants were recruited through the Psychology Department Undergraduate Participant Pool at the University

of Manitoba and participated in exchange for course credit toward their Introduction to Psychology course (Table 1). All participants self-reported having either normal or corrected to normal vision (e.g., wearing glasses, contact lenses, corrective eye-surgery, etc.), and were right-hand dominant, as determined by a modified version of the Edinburgh Handedness Inventory (Oldfield, 1971). All participants also self-reported using their right hand to control the cursor when using a computer. All participants provided informed consent prior to participation, and all procedures were approved by the Psychology/Sociology Research Ethics Board (PSREB) at the University of Manitoba.

Experiment construction

The experiment was built using lab.js (Henninger et al., 2021) a free online study builder designed for the behavioral and cognitive sciences.

Cursor presentation

To ensure accuracy during performance and to avoid any positional biases in cursor position (Phillips et al., 2001, 2003), participants' cursor was set to appear as a "crosshair," rather than the default arrowhead pointer.

Stimuli presentation

The sizes of the stimuli were measured in logical pixels (px), mapped accordingly to the physical pixels of the device's screen based on the device's screen resolution and device-pixel-ratio (DPR). Using logical pixels to design the on-screen

TABLE 1 Demographic information and experimental instructions.

	Demographic information	Instructions
Experiment 1	<i>N</i> = 50 (41 female, 9 male), ages 18–32 years (<i>M</i> = 20.10, <i>SD</i> = 3.38)	"Press the Continue button below to proceed to the next set of trials"
Experiment 2	<i>N</i> = 50 (38 female, 12 male), ages 18–44 years (<i>M</i> = 20.02, <i>SD</i> = 4.40)	"Try to be faster when clicking! Remember: The goal is to click the center of the target circle AS QUICKLY AS POSSIBLE"
Experiment 3	<i>N</i> = 50 (38 female, 11 male, 1 undeclared), ages 17–23 (<i>M</i> = 18.96, <i>SD</i> = 1.59)	"Try to be more accurate when clicking! Remember: The goal is to click the CENTER of the target circle AS ACCURATELY AS POSSIBLE."

stimuli meant the stimuli sizes remained relatively similar across devices; the DPR of devices with significantly higher screen resolutions prevented the stimuli from appearing drastically smaller than on devices with lower resolutions. Unless specified otherwise, the term “pixels” and the abbreviated “px” refers to logical pixels. Information regarding the DPR of the device being used to perform the experiment, as well as the monitor resolution (Table 2), size of window content, and size of the browser viewport was included in the metadata collected by lab.js each time a participant completed the experiment.

The different target types used in this experiment are presented in Table 3. The stimuli were presented within an 800×600 px container, so they could be viewed on a wide range of screen sizes and resolutions. Targets appeared as white circles against a black background, and were presented either alone (Control targets), or surrounded by an annulus of context circles. The size and position of these context circles determined the direction of the illusion. In addition to the traditional variations of the Ebbinghaus illusion (small context circles positioned close to the target vs. large context circles positioned far from the target) a Perceived Large (Far) target was also included (small context circles positioned far from the target), in an attempt to observe the effect of the illusion while controlling for the context circles' proximity to the target.

Experiment hosting

Lab.js supports several options for online study deployment, as well as exportation for offline data collection. The current online study was hosted on Github¹ (Github, 2021), an online, open-source software development platform that provides

internet hosting. Participants were provided with the link to the experiment through the University of Manitoba's online external study management system. As GitHub does not support the saving of participant data, participant data was saved to a secure, realtime database using Google Firebase² (Firebase, 2021) and exported using custom programming in R (R Core Team., 2020)

Procedure

Self-report

Once directed to the experiment website, participants were asked to confirm their use of the touchpad/trackpad of a laptop computer to complete the experiment (use of a physical mouse or touchscreen device to control the on-screen cursor was not permitted). Participants were then presented with a consent form and were required to provide consent before continuing. Next, participants reported to the best of their knowledge the type of device they were using to complete the experiment, as well as the device's screen size, and were asked to confirm once again that they were using their finger on the device's touchpad/trackpad rather than a physical mouse or touchscreen device. Participants then provided demographic information regarding their vision (e.g., normal or corrected-to-normal), sex assigned at birth, and handedness.

Screen set-up

The first task involved participants using their cursor to click on five circular targets (diameter = 2 px) presented in sequence on their computer screen, one target each positioned in the center of the screen, 200 px to the left and right of center (these positions corresponded to the position of the targets during

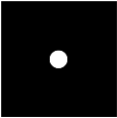
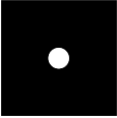
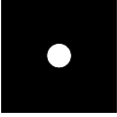
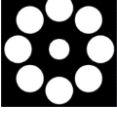
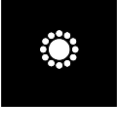

¹ <https://github.com>

² <https://firebase.google.com>

TABLE 2 Screen resolution (logical pixels) and device pixel ratio (DPR).

Experiment 1			Experiment 2			Experiment 3		
Screen resolution	Device pixel ratio (DPR)	Number of participants	Screen resolution	Device pixel ratio (DPR)	Number of participants	Screen resolution	Device pixel ratio (DPR)	Number of participants
1,280 × 720	1.5	6	1,280 × 720	1.5	2	1,280 × 720	1.5	5
1,280 × 800	2	1	1,280 × 800	1	1	1,280 × 800	1	2
1,366 × 768	1	9		2	3		2	2
1,368 × 912	2	1	1,366 × 768	1	12	1,366 × 768	1	9
1,440 × 900	1	5	1,368 × 912	2	1	1,440 × 900	1	8
	2	16	1,440 × 900	1	13		2	14
1,500 × 1,000	2	1		2	14	1,504 × 1,003	1.5	1
1,504 × 1,003	1.5	2	1,536 × 864	1.25	1	1,536 × 864	1.25	6
1,536 × 864	1.25	8	1,600 × 900	1	1	1,792 × 1,120	2	1
1,680 × 1,050	1	1	1,792 × 1,120	2	2	1,920 × 1,080	1	1
						1,920 × 1,200	2	1

TABLE 3 Target type and dimensions.

Target type	Target circle diameter (px)	Context circle diameter (px; proportion of target circle diameter)	Distance from edge of target circle to inner edge of context circle (px)
Control (Small) 	60	–	–
Control (Regular) 	70	–	–
Control (Large) 	80	–	–
Perceived Small 	70	96 (1.37)	58
Perceived Large 	70	27 (0.39)	11
Perceived Large (Far) 	70	27 (0.39)	73

When presented on a 13-inch screen at a resolution of 3000 × 2000 physical pixels and a device pixel ratio = 2, the diameter of the Control Small target measured approximately 11 mm, the diameter of the regular sized targets measured approximately 13 mm, and the diameter of the Control Large target measured approximately 15 mm.

the experimental trials), and 150 pixels above and below the screen’s center. The presentation of each target was preceded by a 200 ms mask to prevent any afterimages of the previous target. The recorded clicks at these target positions were used during analysis to confirm the metadata regarding the device’s screen size and resolution were accurate, as well as to use as a reference point for the target’s position during the experimental trials.

Instructions

Following the screen setup task, participants were presented with a set of instructions explaining the experiment, beginning by asking participants to maintain approximately 2 feet (“2 rulers distance”) between their head and the computer screen, in an attempt to maintain consistent viewing distance across participants (see Li et al., 2020 for additional options for controlling viewing distance). Next, participants were informed that a target circle would appear on the screen and were

instructed to click on the center of the on-screen target “AS QUICKLY AND AS ACCURATELY AS POSSIBLE.” Example images were provided to help describe the task, and to distinguish the “target circle” from the surrounding context circles.

Experimental task

Each trial began with a gray start button (diameter = 30 px), presented 250 px below the center of the screen. Participants were required to click the start button to initiate each experimental trial, and each experimental trial was preceded by a 200 ms mask. Each trial consisted of a target presented either 200 px to the left or right of the screen’s center. Participants completed the trial by moving their cursor to the target and clicking within the target circle’s boundaries, after which the target would disappear, and the start button would reappear to begin the next trial. Only clicking within the target

circle's boundary ended the trial; clicks outside the boundaries were not recorded. There were no time constraints on the presentation of the stimuli, and the target remained on the screen until it was clicked.

Participants completed a set of 12 practice trials, during which each target type was presented twice, once on the left and once on the right side of the screen. Prior to the onset of the experimental trials, participants were once again reminded to “click the center of the target circle as quickly and accurately as possible.” Participants then completed 60 randomized experimental trials (each unique combination of target type and on-screen position shuffled without replacement, then re-shuffled), such that each target type appeared five times on the left side of the screen, and five times on the right. Participants were then given an opportunity to take a break and instructed to “Press the Continue button below to proceed to the next set of trials” before completing another 60 randomized experimental trials.

Perceptual comparisons

After completing the 120 experimental trials, participants completed a forced-choice perceptual size comparison task, in which they were presented with two different target types and were instructed to click which target they believed to be larger. These comparisons were included to check if the illusory context was effectively manipulating the perceived size of the targets. The two targets being compared were never the same type, and each target type was compared with the other five target types twice, appearing once on the left and once on the right side of the screen at positions corresponding to those used during the experimental trials (200 px to the left and right of center) equalling a total of 30 trials. Participants were allowed to take as much time as needed to make their selection. Participants whose responses were incorrect when comparing the three veridically different control targets (Control Small, Control Regular, and Control Large) were excluded from the analysis. After finishing the perceptual comparison task, all participant data was uploaded to the secure database, and participants were debriefed and directed to exit their browser.

Manipulation of task instructions

To test the influence of the experimental instructions on participants' performance, and to explore if the particular demands of the task influenced the degree to which the illusory context affected participants' cursor movements, three separate experiments were conducted. The experimental design and construction were identical for each experiment, with the exception of the message received by participants following the first block of 60 trials. Participants in Experiment 1 were instructed to simply begin the next set of trials, while participants in Experiments 2 and 3 were instructed to prioritize speed or accuracy, respectively, during the next set of trials ([Table 1](#)). To

control for the varying range of devices used by participants, as well as the various screen sizes and resolutions, each experiment was analyzed separately as a within-subject repeated measure design.

Dependent variables

Participants' temporal and spatial cursor data were measured in lab.js using the Mousetrap plugin ([Kieslich and Henninger, 2017](#)). This movement data can be used to explore a wide variety of experimental variables tailored to the specific research question being investigated. In this study, movement variables that would typically be investigated in a traditional reach-to-grasp or reach-to-point study such as accuracy and duration of the movement, deceleration phase, as well as measures of the movement trajectory itself were explored. For each dependent variable, averages were calculated within each unique condition to create a mean condition value for each participant.

Click-point accuracy

The radial error, calculated as the Euclidian distance (px) between the target's center and the location of the participant's click point was used to provide an absolute value representing click-point accuracy. Smaller values indicate click-point positions closer to the target's center and higher accuracy.

Movement time

The amount of time from the onset of cursor movement to the time at which participants clicked the target was measured in milliseconds (ms).

Deceleration phase

To test for an effect of the illusion on participants' online control of the cursor, the phase of deceleration—defined as the proportion of the overall (100%) movement following the point in time at which peak velocity was reached—was analyzed.

Area under the curve

Cursor trajectories were spatially normalized using 101 equidistant points (i.e., 0–100% of the movement distance) along the original cursor trajectory using the `mt_spatialize` function in the `mousetrap` package. The Area under the curve (AUC) was defined as the geometric area (px) between the trajectory and an idealized (straight) path connecting the trajectory's start and end positions. The R function `polysimplify` from the `polyclip` package ([Johnson and Baddeley, 2019](#)) was used to separate the cursor deviations from the idealized path and the `polyarea` function from the `pracma` package ([Borchers, 2021](#)) was used to combine the area of these deviations. Doing so produced an absolute deviation value by treating participants' deviations

as additive rather than subtractive (the default method in the mousetrap package). Larger AUC values represent greater deviation from the idealized path, and a more curved cursor trajectory.

Time of maximum deviation

The stage of movement—defined as a proportion of the overall (100%) movement—at which the cursor position deviated the farthest from the idealized path connecting the trajectory's start and end positions.

Number of directional changes

Using the same spatially normalized trajectories mentioned above, the frequency at which participants changed the direction of their cursor movement in either the horizontal or vertical axes during their movement toward the target in each trial was counted and averaged to create a mean condition value for each participant.

Data analysis

The cursor data collected by the Mousetrap plugin was analyzed using the Mousetrap package (Kieslich et al., 2019) in R (R Core Team., 2020). Each trial ended once participants clicked the target, and therefore the final x and y coordinates of these cursor trajectories were used to define the position of the participants' click point on each trial. Each cursor trajectory was inspected manually, to ensure the cursor position was recorded effectively throughout the trial. Additionally, as the logging resolution (the intervals at which the cursor position was recorded throughout the movement) had the potential to vary across devices, the logging resolution of each dataset was checked using the `mt_check_resolution` function in the mousetrap package. In all cases, the logging resolution was deemed satisfactory.

Experimental trials involving the Perceived Large (Far) target were removed from the analysis (see first paragraph of the "Results" section below). Each dependent variable was therefore analyzed using a 2 (Time: Pre-Break vs. Post-Break) \times 2 (Position: Left vs. Right) \times 5 (Target Type: Control Small vs. Perceived Small vs. Control Regular vs. Perceived Large vs. Control Large) within-subjects repeated measures ANOVA. All statistical analyses were conducted using SPSS (version 23.0). A Greenhouse-Geisser correction was used to address any violations to sphericity. Violations to the assumption of normality were identified by inspecting the normality of the residual values produced by the repeated measures ANOVA. In cases where the residual values were significantly and consistently non-normal, a transformation was applied to correct the non-normal data. All analyses were conducted using $\alpha = 0.05$, and Bonferroni adjusted p -values were applied

to all *post-hoc* comparisons used to analyze any significant interactions.

Results

In all three experiments, analysis of participants' perceptual comparison scores consistently indicated that the Perceived Large (Far) target was not successful in inducing the desired increase in perceived target size (participants reported an increase in the target's perceived size in as few as 32% and no more than 58% of comparisons). This is likely due to the increased distance between the context circles and the target circle. Proximity of the context circles to the target circle is known to play an important role in the direction and magnitude of the Ebbinghaus illusion's effect, with closer context circles increasing the size of the target circle, and farther context circles minimizing the size of the target circle (Massaro and Anderson, 1971; Knol et al., 2015), as also occurs in the Delboeuf illusion (Roberts et al., 2005). Based on the lack of any useful effect of the illusion, experimental trials involving the Perceived Large (Far) target were not analyzed.

Experiment 1

Excluded data

A coding error made it possible for participants to begin their cursor movements immediately after clicking the start button, during the 200 ms mask prior to presentation of the target. This meant that any cursor movement that was executed during the 200 ms mask was not captured as part of the experimental trial. In total, 1.84% of all trials involved uncaptured cursor movement during the 200 ms mask and were therefore excluded from analysis. An additional 0.10% of all trials were removed due to missing time timestamp data (the timepoints throughout the trial at which cursor position was captured). Trials lasting longer than 5,000 ms to perform the task were also removed. This cut-off was determined to be excessive based on inspection of participants' movement time data during analysis and accounted for 0.16% of the total number of trials. Finally, while the onscreen target represented the only "clickable area" on the screen, this clickable area was defined using square boundaries, which meant that in rare cases, participants could in fact click "outside" the circular target, in the corners of the square boundary. This occurred in 0.02% of trials, all of which were excluded from analysis. In total, 2.12% of experimental trials were excluded from analysis in Experiment 1.

TABLE 4 Average perceptual comparison durations (ms).

Target position		Average duration (ms)		
Left	Right	Experiment 1	Experiment 2	Experiment 3
Control Small	Perceived Small	2513.83 (1499.19)	2251.92 (1975.41)	2737.29 (2196.80)
Control Small	Control Regular	1570.12 (611.02)	1387.04 (476.81)	1717.00 (654.22)
Control Small	Perceived Large (Far)	1925.68 (881.48)	1902.26 (1278.20)	2305.79 (1218.07)
Control Small	Perceived Large	1598.53 (519.55)	1618.43 (712.68)	1994.65 (948.93)
Control Small	Control Large	1412.72 (541.48)	1268.44 (401.80)	1573.73 (588.57)
Perceived Small	Control Small	2833.39 (1958.59)	2343.64 (1814.58)	3043.41 (1937.97)
Perceived Small	Control Regular	1827.62 (604.81)	1653.86 (629.30)	2424.98 (1856.95)
Perceived Small	Perceived Large (Far)	2609.35 (4426.98)	2467.01 (1563.89)	2473.24 (1268.47)
Perceived Small	Perceived Large	1908.66 (722.92)	2068.09 (1600.03)	2451.17 (1894.03)
Perceived Small	Control Large	1682.91 (519.62)	1561.28 (542.36)	1763.56 (811.28)
Control Regular	Control Small	1652.10 (470.85)	1387.32 (441.16)	1715.82 (600.22)
Control Regular	Perceived Small	2231.34 (1175.06)	2077.99 (1575.86)	2142.87 (928.95)
Control Regular	Perceived Large (Far)	3514.77 (6342.86)	2754.90 (4035.91)	2938.42 (1957.30)
Control Regular	Perceived Large	2197.54 (1705.32)	1886.28 (1089.75)	2381.06 (1510.89)
Control Regular	Control Large	1608.49 (510.66)	1481.97 (537.51)	1830.50 (1154.93)
Perceived Large (Far)	Control Small	2012.69 (1015.28)	1640.08 (802.74)	1861.31 (642.35)
Perceived Large (Far)	Perceived Small	2497.80 (2788.96)	2314.77 (2074.66)	2416.01 (1445.88)
Perceived Large (Far)	Control Regular	2478.00 (1622.94)	2918.56 (5145.41)	2813.16 (1978.17)
Perceived Large (Far)	Perceived Large	2267.54 (1174.96)	3072.69 (5504.96)	3148.43 (2219.74)
Perceived Large (Far)	Control Large	1694.13 (589.59)	1683.71 (967.56)	1840.67 (676.94)
Perceived Large	Control Small	1833.64 (607.44)	1542.29 (592.46)	1929.27 (794.80)
Perceived Large	Perceived Small	2097.17 (988.34)	1888.54 (982.18)	1996.51 (1172.08)
Perceived Large	Control Regular	2255.42 (1028.70)	2268.09 (2099.42)	2642.57 (1732.29)
Perceived Large	Perceived Large (Far)	2483.62 (1630.25)	2532.55 (3727.42)	3149.33 (2248.89)
Perceived Large	Control Large	2170.40 (1606.41)	1695.23 (759.18)	2131.64 (1061.51)
Control Large	Control Small	1533.86 (463.12)	1258.86 (394.20)	1579.30 (592.23)
Control Large	Perceived Small	1841.06 (968.74)	1531.24 (580.31)	1853.19 (604.94)
Control Large	Control Regular	1582.56 (570.45)	1380.62 (362.76)	1910.71 (1120.91)
Control Large	Perceived Large (Far)	1763.81 (626.15)	1652.93 (762.92)	1847.74 (718.73)
Control Large	Perceived Large	2349.27 (1549.00)	2198.07 (1378.20)	2325.61 (1051.54)

Standard deviations presented in parentheses.

Perceptual comparisons

As the time spent comparing the onscreen targets may have influenced participants' perceptual responses, the average trial durations for each perceptual comparison are provided for context in **Table 4**. Participants' perceptual comparison scores are provided in **Table 5**. Participants' responses followed the direction of the illusion with a generally high consistency: over 75%, except for the comparisons involving the Perceived Large (Far) target. Additionally, a small portion of participants reported the Perceived Small target as being smaller than the veridically smaller Control Small target (20% when the Perceived Small target was on the left side of the screen, and 22% when it was on the right side of the screen). When comparing the Perceived Large target on the right side of the screen with the Control Large target on the left side of the screen, 25% of participants reported the Perceived Large target

as being larger, however, this pronounced effect of the illusion disappeared when the target positions were reversed (0% when the Perceived Large target was on the left and the Control Large target was on the right).

Click-point accuracy

Examining the distributions of participants' average accuracy scores within each condition indicated non-normal, moderately to severely positively skewed data in all conditions. To address this violation to normality, a log transformation was applied to the data. The data reported here have been back-transformed into their original units for ease of interpretation.

A significant main effect of Time, $F(1, 49) = 7.09$, $p < 0.05$, $\eta_p^2 = 0.13$, indicated that participants were

TABLE 5 Experiment 1 perceptual comparisons.

Onscreen position		Right					
		Control Small	Perceived Small	Control	Perceived Large (Far)	Perceived Large	Control Large
Left	Control Small	–	80%*	–	90%	100%	–
	Perceived Small	78%*	–	92%	92%	94%	96%
	Control	–	88%	–	44%	76%	–
	Perceived Large (Far)	94%	94%	32%	–	82%	98%
	Perceived Large	100%	96%	76%	84%	–	100%
	Control Large	–	98%	–	94%	76%*	–

Scores represent the percent of comparisons that demonstrated the expected size ordering (Smallest to Largest): Control (Small) < Perceived Small < Control < Perceived Large (Far) < Perceived Large < Control Large. **Bolded scores represent the comparisons between the same-sized targets.** Comparisons between Control stimuli (all 100%) not included.

*The fact that this value is less than 100% suggests that on the remaining percent of trials participants reported the illusory target as smaller or larger than the veridically smaller (Control Small) or larger (Control Large) targets respectively, suggesting the presence of an exaggerated illusory effect.

more accurate in their click positions during the first block of trials ($M = 3.62$ px, 95% CI [3.04, 4.31]) compared to the second block of trials ($M = 3.94$ px, 95% CI [3.27, 4.73]). A significant main effect of Target, $F(4, 196) = 10.91$, $p < 0.001$, $\eta_p^2 = 0.18$, indicated that participants were generally most accurate when clicking on the Control Small target (Figure 1). This is likely due to the Control Small target representing a smaller clickable area compared to the other targets, requiring participants to be more accurate overall. However, there were no significant comparisons amongst the three same-sized targets (Perceived Small, Perceived Large and Control Regular targets), suggesting the presence of the illusion did not influence participants' clicking accuracy.

Movement time

A significant main effect of Target, $F(4, 196) = 4.87$, $p < 0.01$, $\eta_p^2 = 0.09$, suggested the type of target influenced participants' speed when performing the task (Figure 2). However, the only significant comparison was between the Perceived Large and the Control Large targets; movement time was significantly longer when clicking on the Perceived Large target.

Deceleration phase

The distributions of the average deceleration phase lengths were determined to be non-normal (moderately to severely negatively skewed). A square transformation was applied to the data to correct this violation of normality. The data reported here have been back-transformed into their original units for ease of interpretation.

Participants' deceleration phases were not affected by Time, Target Type, or Position (all $ps > 0.05$). The average deceleration period across all conditions was 87.28% ($SD = 0.62\%$) of the total movement.

Area under the curve

Figure 3 presents an example of the cursor trajectories executed by one participant when clicking on the Perceived Large (Figure 3A) and Perceived Small target (Figure 3B). A significant main effect of Position, $F(1, 49) = 7.01$, $p < 0.05$, $\eta_p^2 = 0.13$, indicated that participants executed more curved cursor movements when the target was presented on the right side of the screen ($M = 17309.72$ px, 95% CI [15342.79, 19276.64]) compared to when presented on the left ($M = 14608.33$ px, 95% CI [12790.36, 16426.31]).

Time of maximum deviation

A main effect of Target was found to be significant, $F(4, 196) = 3.331$, $p < 0.05$, $\eta_p^2 = 0.06$, and a general trend suggested the maximum deviation occurred earliest when clicking on the Perceived Large target ($M = 17.83\%$ of the movement, 95% CI [16.18, 19.49]), followed by the Control Regular ($M = 18.08\%$, 95% CI [16.56, 19.60]), Control Small ($M = 18.43\%$, 95% CI [16.73, 20.13]), Control Large ($M = 18.92\%$, 95% CI [17.35, 20.49]) and Perceived Small ($M = 19.06\%$, 95% CI [17.56, 20.57]) targets, however, all comparisons were non-significant (all $ps > 0.05$).

Number of directional changes

A significant main effect of Target, $F(4, 196) = 3.91$, $p < 0.01$, $\eta_p^2 = 0.07$ (Figure 4), showed that participants made significantly more directional changes when clicking the Perceived Large target in comparison to the Perceived Small and Control Large targets. There were no significant differences in the number of directional changes between any of the other target types. An increased number of directional changes were also made when the target was positioned on the right side of the screen ($M = 2.51$, 95% CI [2.29, 2.73]) compared to when positioned on the left side of the screen ($M = 2.28$, 95% CI [2.05, 2.51]), as indicated by a

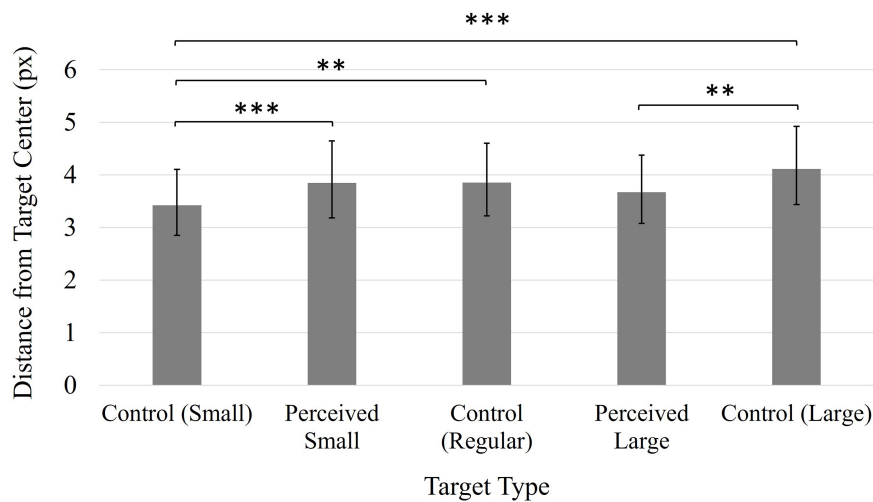


FIGURE 1

Average distance from click position to target center. Values have been back-transformed into original measurement value (px). Smaller values indicate higher accuracy. Error bars represent 95% confidence intervals. ** $p < 0.01$, *** $p < 0.001$.

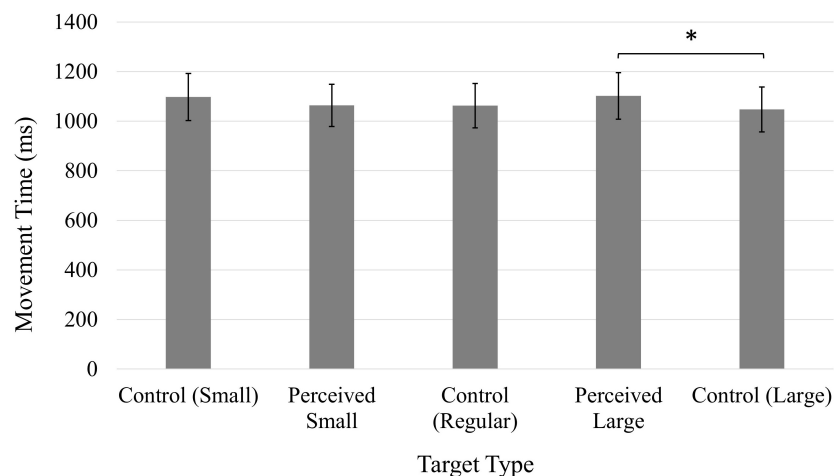


FIGURE 2

Average movement time (ms). Error bars represent 95% confidence intervals. * $p < 0.05$.

significant main effect of Position, $F(1, 49) = 8.54$, $p < 0.01$, $\eta_p^2 = 0.15$.

Experiment 2

Excluded data

In total, 2.16% of all trials were excluded from analysis in Experiment 2 (early cursor movement during the 200 ms mask: 1.64%, unusable cursor/timestamp data: 0.38%, trial duration longer than 5,000 ms: 0.10%, click-point outside target boundaries: 0.04%).

Perceptual comparisons

Participants' perceptual comparison scores are provided in Table 6. As in Experiment 1, participants' perceptual comparison scores indicated the illusory context successfully influenced participants' perceptions of target size, except for the Perceived Large (Far) target, which once again had comparatively low scores. Again, the illusion appeared to have an exaggerated effect in a small subset of responses.

Click-point accuracy

As in Experiment 1, participants' average accuracy scores violated the assumption of normality (distributions ranged from moderately to severely positively skewed on a consistent basis).

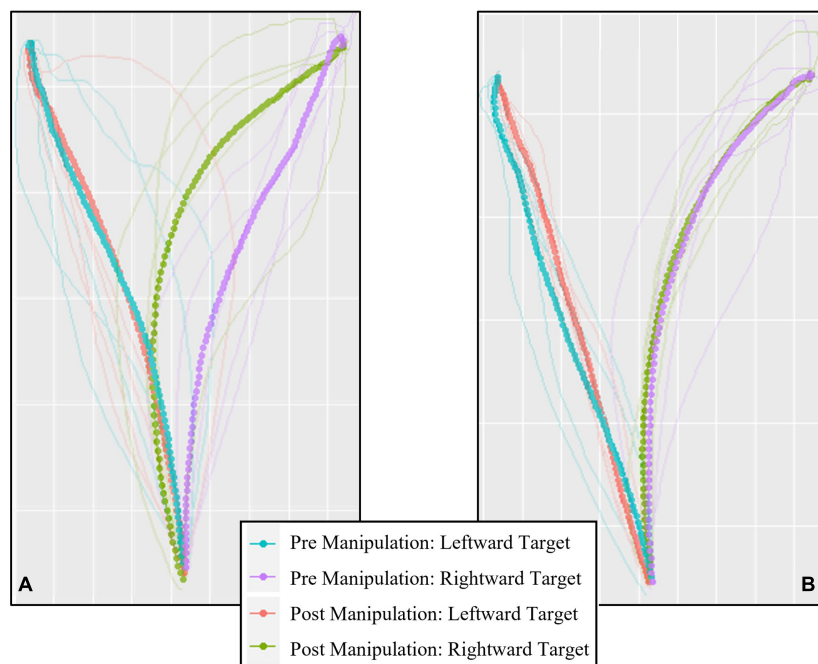


FIGURE 3

An example of the trajectories generated by a single participant executing cursor movements toward the Perceived Large (A) and Perceived Small (B) target. Solid lines represent data from each trial, dotted lines represent the average trajectory for each condition. Plot generated using the `mt_plot_aggregate` function from the `mousetrap` package in R (Kieslich and Henninger, 2017).

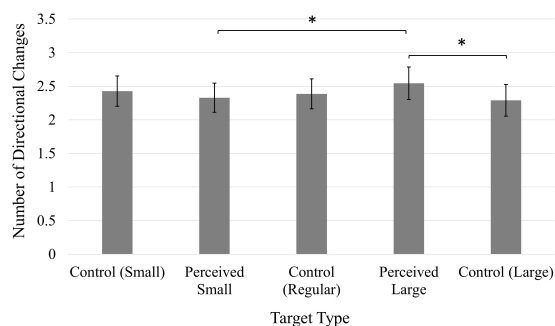


FIGURE 4

Average number of directional changes. Error bars represent 95% confidence intervals. * $p < 0.05$.

A log transformation was applied to the data, and the values reported here have been back-transformed into the original units (px). A significant main effect of Time, $F(1, 49) = 33.47$, $p < 0.001$, $\eta_p^2 = 0.41$, showed that participants were less accurate in the second block of trials following the manipulation ($M = 6.40$ px, 95% CI [5.15, 7.96]) compared to the first, pre-manipulation block ($M = 4.48$ px, 95% CI [3.64, 5.50]). A significant main effect of Position, $F(1, 49) = 6.92$, $p < 0.05$, $\eta_p^2 = 0.12$, also indicated that participants were more accurate when the target was presented on the right side of the screen

($M = 5.21$ px, 95% CI [4.23, 6.44]) than when presented on the left ($M = 5.50$ px, 95% CI [4.51, 6.68]).

Finally, a main effect of Target, $F(4, 196) = 6.92$, $p < 0.001$, $\eta_p^2 = 0.12$ (Figure 5), was also found to be significant. Accuracy was generally worse when participants clicked on the Control Large target. There were no significant comparisons amongst the same-sized targets (Perceived Small, Perceived Large, and Control Regular).

Movement time

Participants' average movement time scores were consistently non-normal (moderately to severely positively skewed), and a log transformation was applied to the data. The values reported here have been back-transformed into the original units (ms). Participants were significantly faster during the second block of trials following the experimental manipulation ($M = 820.35$ ms, 95% CI [741.31, 905.73]) compared to the first block of trials ($M = 968.28$ ms, 95% CI [874.98, 1069.06]), as confirmed by a significant main effect of Time, $F(1, 49) = 32.99$, $p < 0.001$, $\eta_p^2 = 0.40$. A main effect of Position, $F(1, 49) = 9.74$, $p < 0.01$, $\eta_p^2 = 0.17$, was also significant, and participants were faster when the target was presented on the left side of the screen ($M = 877.00$ ms, 95% CI [796.16, 968.28]) compared to targets presented on the right ($M = 903.65$ ms, 95% CI [822.25, 993.12]). A significant main effect of Target, $F(4, 196) = 6.63$, $p < 0.001$, $\eta_p^2 = 0.12$

TABLE 6 Experiment 2 perceptual comparisons.

Onscreen position		Right					
		Control Small	Perceived Small	Control	Perceived Large (Far)	Perceived Large	Control Large
Left	Control Small	–	84%*	–	96%	96%	–
	Perceived Small	72%*	–	84%	82%	94%	98%
	Control	–	92%	–	44%	78%	–
	Perceived Large (Far)	98%	96%	44%	–	76%	90%
	Perceived Large	98%	96%	88%	80%	–	82%*
	Control Large	–	100%	–	96%	82%*	–

Scores represent the percent of comparisons that demonstrated the expected size ordering (Smallest to Largest): Control (Small) < Perceived Small < Control < Perceived Large (Far) < Perceived Large < Control Large. **Bolded scores represent the comparisons between the same-sized targets.** Comparisons between Control stimuli (all 100%) not included. *The fact that this value is less than 100% suggests that on the remaining percent of trials participants reported the illusory target as smaller or larger than the veridically smaller (Control Small) or larger (Control Large) targets respectively, suggesting the presence of an exaggerated illusory effect.

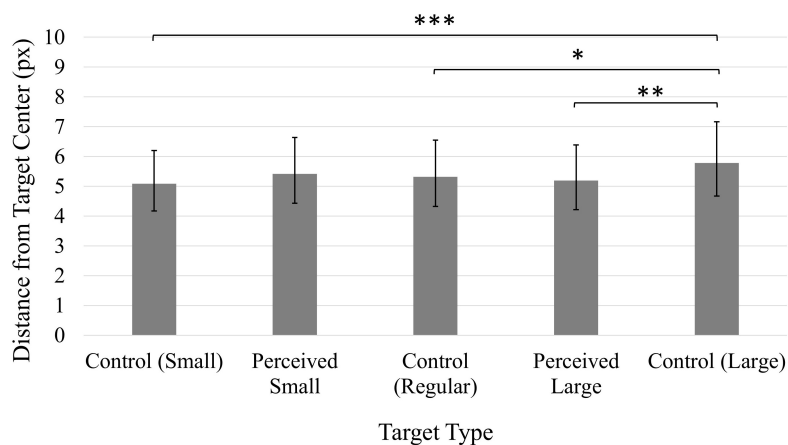


FIGURE 5

Average distance from click position to target center. Values have been back-transformed into original measurement value (px). Smaller values indicate higher accuracy. Error bars represent 95% confidence intervals. * $p < 0.01$, ** $p < 0.01$, *** $p < 0.001$.

(Figure 6), indicated the decreased accuracy observed when clicking on the Control Large target was also associated with shorter movement times; participants were faster when clicking on the Control Large target compared to the Control Small, Perceived Small, and Perceived Large targets.

Deceleration phase

Deceleration phases were significantly longer during the first block of trials ($M = 86.56\%$ of total movement, 95% CI [84.80, 88.314]) compared to the second block of trials ($M = 83.50\%$ of total movement, 95% CI [81.60, 85.40]), as indicated by a significant main effect of Time, $F(1, 49) = 31.07$, $p < 0.001$, $\eta_p^2 = 0.39$. A significant main effect of Target, $F(2, 196) = 3.81$, $p < 0.01$, $\eta_p^2 = 0.07$, indicated longer deceleration phases when clicking on the Perceived Small target ($M = 85.95\%$, 95% CI [84.16, 87.73]) compared to the Control

Large target ($M = 84.02\%$, 95% CI [81.98, 86.07]; $p < 0.05$). There were no other significant comparisons involving the Control Small ($M = 84.68\%$, 95% CI [82.80, 86.56]), Control Regular ($M = 85.16\%$, 95% CI [83.51, 86.81]) and Perceived Large ($M = 85.33\%$, 95% CI [83.38, 87.28]) targets (all $ps > 0.05$).

Area under the curve

Cursor trajectories were significantly more curved during the second block of trials, after the manipulation ($M = 17489.86$ px, 95% CI [15834.25, 19145.46]) than compared to the first block of trials ($M = 16131.00$ px, 95% CI [14476.78, 17785.23]), as confirmed by a significant main effect of Time, $F(1, 49) = 7.08$, $p < 0.05$, $\eta_p^2 = 0.13$. A significant main effect of Position, $F(1, 49) = 25.07$, $p < 0.001$, $\eta_p^2 = 0.34$, indicated that cursor trajectories were also more curved when the target was presented on the right side of the screen ($M = 18914.69$ px,

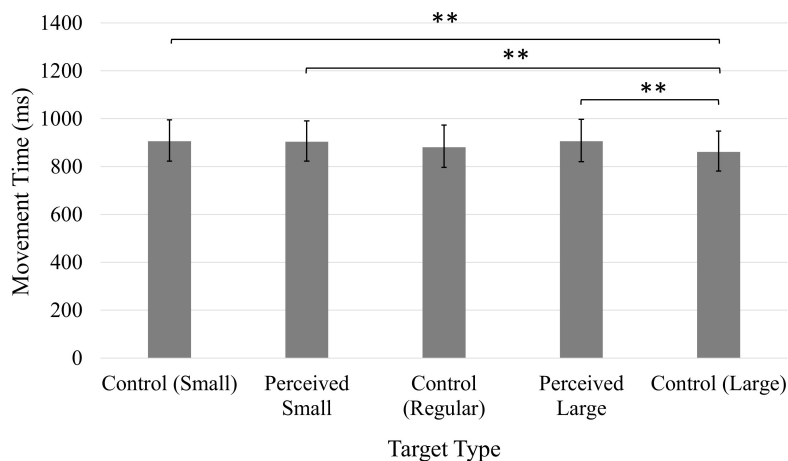


FIGURE 6

Average movement time (ms). Error bars represent 95% confidence intervals. ** $p < 0.01$.

95% CI [16817.29, 21012.08]) than when presented on the left ($M = 14706.17$ px, 95% CI [13299.45, 16112.90]).

Time of maximum deviation

A significant Time \times Position interaction, $F(1, 49) = 6.759$, $p < 0.05$, $\eta_p^2 = 0.12$, indicated that while the maximum deviation occurred earlier during the first block of trials for both leftward presented targets (Pre Manipulation: $M = 20.43\%$ of movement time, 95% CI [18.46, 22.39], Post Manipulation: $M = 23.27\%$, 95% CI [21.30, 25.23], $p < 0.001$) and rightward presented targets (Pre Manipulation: $M = 18.58\%$, 95% CI [16.67, 20.49], Post Manipulation: $M = 22.93\%$, 95% CI [20.87, 24.99], $p < 0.001$), differences in the time of maximum deviation between left- and rightward presented targets only occurred during the first block of trials, during which the maximum deviation occurred earlier for rightward positioned targets ($p < 0.001$) than rightward positioned targets. There was no significant difference in the time at which the maximum deviation occurred between left- and rightward presented targets in the second block of trials ($p > 0.05$).

Number of directional changes

A significant Position \times Stimuli interaction, $F(4, 196) = 2.76$, $p < 0.05$, $\eta_p^2 = 0.05$ (Figure 7), indicated that an increased number of directional changes were made when clicking on each target type when positioned on the right side of the screen compared to when presented on the left side, except for the Perceived Small target, for which the number of directional changes did not significantly differ between target positions ($p > 0.05$). There were no significant differences in the number of directional changes made by participants when clicking on targets presented on the left side of the screen (all

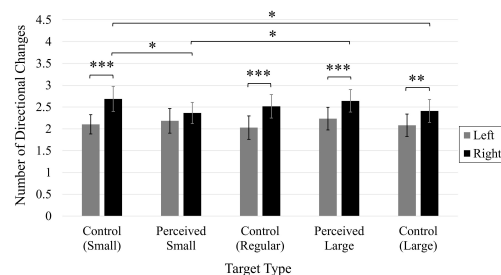


FIGURE 7

Average number of directional changes. Error bars represent 95% confidence intervals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

$ps > 0.05$). However, significantly more corrections were made when clicking on the Control Small target in comparison to the Perceived Small target, and the Control Large target when these targets were presented on the right side of the screen. Cursor movements toward the Perceived Large target also involved more corrections in comparison to the Perceived Small target when presented on the right side of the screen. The fact that these comparisons were only significant when the targets were presented on the right side of the screen suggests the effect of the illusion may have been highlighted by the increased difficulty associated with a rightward movement on the trackpad.

Experiment 3

Excluded data

In total, 3.14% of all trials were excluded from analysis in Experiment 3 (early cursor movement during the 200 ms mask:

2.60%, unusable cursor/timestamp data: 0.10%, trial duration longer than 5,000 ms: 0.44%).

Perceptual comparisons

Participants' perceptual comparison scores are provided in [Table 7](#). As in Experiments 1 and 2, the illusory context successfully influenced participants' perceptions of target size, however, this was not the case for the Perceived Large (Far) target. An exaggerated influence of the illusion was observed in a small portion of participants' responses, regardless of the target's on-screen position.

Click-point accuracy

As was the case for Experiments 1 and 2, participants' average accuracy scores were consistently positively skewed, and a log transformation was applied to the data. The values reported here have been back-transformed into the original units (px). A significant main effect of Time, $F(1, 49) = 42.44$, $p < 0.001$, $\eta_p^2 = 0.46$, indicated that participants' accuracy increased following the manipulation ($M = 2.37$, 95% CI [2.05, 2.74]) compared to before the manipulation ($M = 4.27$, 95% CI [3.44, 5.29]). A significant main effect of Target, $F(4, 196) = 5.66$, $p < 0.001$, $\eta_p^2 = 0.10$ ([Figure 8](#)), indicated the Control Large target generated significantly worse accuracy in comparison to the Control Small and Perceived Large targets. All other comparisons were non-significant ($ps > 0.05$).

Movement time

A significant main effect of Time, $F(1, 49) = 61.47$, $p < 0.001$, $\eta_p^2 = 0.56$, indicated that participants were slower following the manipulation ($M = 1358.33$ ms, 95% CI [1246.81, 1469.86]) compared to before the manipulation ($M = 1059.90$ ms, 95% CI [951.28, 1168.52]).

Deceleration phase

Deceleration phases were significantly longer in the second block of trials, following the manipulation ($M = 89.03\%$, 95% CI [87.64, 90.43]) compared to the first block of trials ($M = 85.86\%$, 95% CI [83.99, 87.73]), as indicated by a significant main effect of Time, $F(1, 49) = 22.45$, $p < 0.001$, $\eta_p^2 = 0.31$. A significant main effect of Target was also observed, $F(4, 196) = 3.298$, $p < 0.05$, $\eta_p^2 = 0.06$, however there were no significant comparisons between the different target types (all $ps > 0.05$); Control Small ($M = 87.14\%$, 95% CI [85.56, 88.72]), Perceived Small ($M = 87.88\%$, 95% CI [86.24, 89.53]), Control Regular ($M = 87.09\%$, 95% CI [85.25, 88.92]), Perceived Large ($M = 88.35\%$, 95% CI [86.88, 89.82]), Control Large ($M = 86.77\%$, 95% CI [85.12, 88.41]).

Area under the curve

A three-way Time \times Position \times Target interaction was shown to be significant, $F(4, 196) = 2.85$, $p < 0.05$, $\eta_p^2 = 0.06$ ([Figure 9](#)). Prior to the manipulation, trajectories were more

curved when the Perceived Small, Perceived Large, and Control Large targets were presented on the right side of the screen compared to the left side; the position of the target had no influence on the Control Small and Control Regular targets ($ps > 0.05$). After the manipulation, trajectories were more curved when the Control Small, Perceived Small, and Control Regular targets were presented on the right side of the screen compared to the left side; the position of the target had no influence on the Perceived Large and Control Large targets ($ps > 0.05$). When clicking on Control Small and Control Regular targets presented on the right side of the screen, trajectory curvature increased following the manipulation. Otherwise, the manipulation did not influence trajectory curvature ($ps > 0.05$). There were no significant comparisons between any of the target types on either the left or right side of the screen, or before or after the manipulation ($ps > 0.05$).

Time of maximum deviation

A significant main effect of Time, $F(1, 49) = 30.03$, $p < 0.001$, $\eta_p^2 = 0.38$ indicated the maximum deviation occurred earlier in the second block of trials, following the manipulation ($M = 15.41\%$ of the total movement, 95% CI [13.89, 16.92]) compared to the first block of trials ($M = 19.71\%$, 95% CI [17.69, 21.74]).

Number of directional changes

A significant three-way Time \times Position \times Target interaction was observed, $F(4, 196) = 4.75$, $p < 0.01$, $\eta_p^2 = 0.09$ ([Figure 10](#)). Prior to the manipulation, participants made significantly more directional changes when clicking on rightward positioned Perceived Small and Control Regular targets in comparison to when these targets were presented on the left side of the screen. After the manipulation, more directional changes were observed when each target was presented on the right side of the screen compared to the left side, with the exception of the Control Large target, for which the number of directional changes did not differ between onscreen positions ($p > 0.05$). The number of directional changes increased following the manipulation when clicking on rightward positioned Control Small, Control Regular, and Perceived Large targets. The number of directional changes also increased post-manipulation when clicking on Control Large targets presented on the left side of the screen. Prior to the manipulation, the number of directional changes did not significantly differ between target types (all $ps > 0.05$). Following the manipulation, however, participants made significantly more directional changes when clicking on the Control Small target compared to the Perceived Large target when these targets were presented on the left side of the screen. Otherwise, there were no significant comparisons between the different targets on either the left or right side of the screen, or before or after the manipulation ($ps > 0.05$).

TABLE 7 Experiment 3 perceptual comparisons.

Onscreen position		Right					
		Control Small	Perceived Small	Control	Perceived Large (Far)	Perceived Large	Control Large
Left	Control Small	–	80%*	–	100%	100%	–
	Perceived Small	76%*	–	88%	86%	98%	96%
	Control	–	84%	–	46%	82%	–
	Perceived Large (Far)	98%	94%	58%	–	80%	92%
	Perceived Large	100%	92%	76%	76%	–	86%*
	Control Large	–	94%	–	94%	84%*	–

Scores represent the percent of comparisons that demonstrated the expected size ordering (Smallest to Largest): Control (Small) < Perceived Small < Control < Perceived Large (Far) < Perceived Large < Control Large. **Bolded scores represent the comparisons between the same-sized targets.** Comparisons between Control stimuli (all 100%) not included. *The fact that this value is less than 100% suggests that on the remaining percent of trials, participants reported the illusory target as smaller or larger than the veridically smaller (Control Small) or larger (Control Large) targets respectively, suggesting the presence of an exaggerated illusory effect.

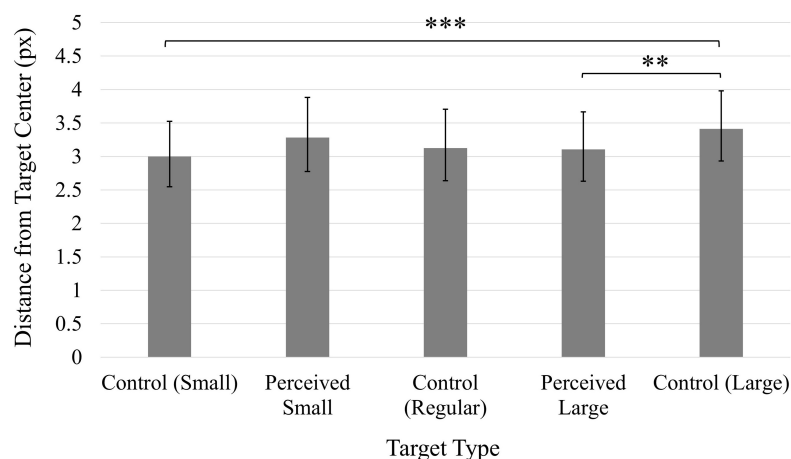


FIGURE 8

Average distance from click position to target center. Values have been back-transformed into original measurement value (px). Smaller values indicate higher accuracy. Error bars represent 95% confidence intervals. ** $p < 0.01$, *** $p < 0.001$.

General discussion

The goal of this study was to test the feasibility of using remote data collection methods to study the type of visually guided actions typically measured within the laboratory setting. Remote data collection provides several potential advantages for behavioral research, including the opportunity to measure a large sample size in a short period of time, as well as presenting a cost-effective and convenient option for both the experimenter and the participant, who can participate in the comfort of their own home. The methods outlined in this study present a novel approach to the collection and analysis of visually guided cursor movement and present an example of how these measures can be applied to investigations of visual perception and action.

Participants' perceptual comparison scores consistently indicated that the presence of the illusion successfully influenced the perceived size of the targets. However, all three experiments failed to demonstrate an influence of perceived target size on click-point accuracy or movement time. In

this sense, these results seem to provide evidence in favor of a visually guided action system that operates separate from the influence of perception, at least in the context of the visuomotor transformation used in this study (i.e., transformation of the proximal digit movement to the distal cursor movement). The results of the current study are similar to those of a study conducted by Janczyk et al. (2013) in which participants' perceptual judgments were influenced by irrelevant stimulus dimensions during a Garner-interference speed classification task, while cursor movements directed toward these stimuli were unaffected. Thus, there appears to be increasing evidence that the cursor movements are unaffected by perceptual intrusions.

While the illusory context did not appear to influence participants' click-point accuracy or movement time as expected, these results suggest the illusory context did influence the trajectories of participants' cursor movements toward the target. More directional changes were observed when clicking the Perceived Large target in comparison to the Perceived Small

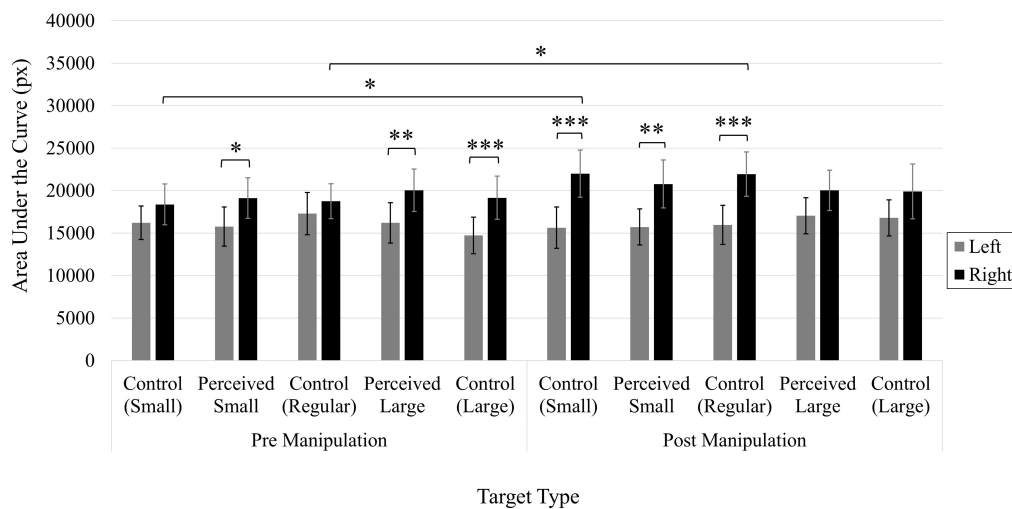


FIGURE 9

Average area under the curve (px). Error bars represent 95% confidence intervals. ** $p < 0.01$, *** $p < 0.001$.

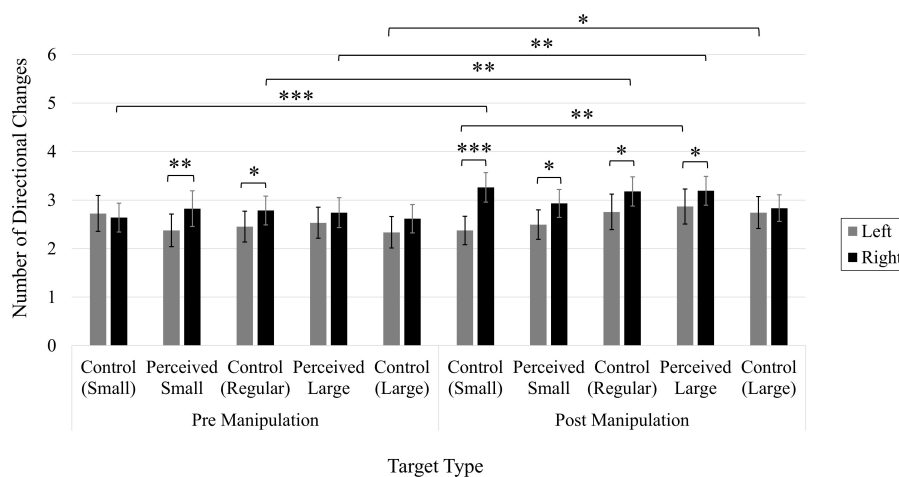


FIGURE 10

Average number of directional changes. Error bars represent 95% confidence intervals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

target in Experiments 1 and 2. In general, cursor movements directed toward rightward positioned targets demonstrated larger AUCs (more curved trajectories) and more directional changes, suggesting participants may have had more difficulty with rightward movements on the trackpad compared to leftward movements. However, the Perceived Small target was the only target type to not produce an increase in the number of directional changes when positioned on the right compared to the left side of the screen in Experiment 2. This left-right differentiation was observed in the trajectory curvature and number of directional changes toward the Perceived Small target in Experiment 3, however, the increased emphasis on accuracy in this experiment did not influence the Perceived Small target as it did the other target types. Taken together, these results

suggest participants' cursor paths were not influenced to the same degree by the on-screen position of the target or the demands of the task when clicking on the Perceived Small target in comparison to the other target types.

Whereas the perception-action model predicts participants' cursor paths will be unaffected by the illusory context, and therefore should remain similar regardless of the type of target, a planning-control model (e.g., Glover and Dixon, 2002; Glover, 2004) might predict the number of corrections to vary as a result of the illusory context, as was observed in this study. Specifically, if the illusory context decreased the perceived size of the target (Perceived Small target) during the planning stage, the control stage of that movement may require less corrections than that of a movement that was planned toward

a target perceived to be larger than its physical size (Perceived Large target). This could potentially explain the increased number of corrections and resulting increased accuracy and movement time observed when clicking the Perceived Large target. Further, if cursor movements toward the Perceived Small target benefited from a more accurate planning stage, these movements would likely be less influenced by the present task demands (i.e., target position or speed and accuracy of the movement) than less accurately planned movements, as observed in Experiments 2 and 3. Further evidence in favor of this explanation comes from the time at which the maximum deviation from the idealized trajectory occurred. A significant effect of Target Type in Experiment 1 suggested the maximum deviation occurred earliest when clicking the Perceived Large target, and latest when clicking the Perceived Small target. The early deviations in response to the Perceived Large target may represent a less accurate planning stage and thus explain why more corrective movements were required when clicking this type of target compared to the Perceived Small target, for which the maximum deviation occurred later, possibly indicating a more accurate planning stage. However, despite the significant main effect, the more stringent Bonferroni corrected pairwise comparisons between the different target types did not reach statistical significance.

A planning-control model of this nature such as the one proposed by [Glover and Dixon \(2002\)](#) and [Glover \(2004\)](#) is not without its own share of criticisms, however ([Danckert et al., 2002](#); [Franz, 2003](#); [Goodale and Milner, 2004](#); [Handlovsky et al., 2004](#); [Westwood, 2004](#); [Franz et al., 2005](#)). Nevertheless, the results of this study appear to suggest the planning of these cursor movements may have been influenced by the illusory context, while end-point measures such as final accuracy and movement time were not. As such, the results of the current study may be explained by both the perception-action and planning-control models of visually guided action in the control of onscreen cursor movements.

Alternatively, it is also possible that participants' cursor movements were not influenced by the effect of the illusion, but rather these differences were produced by the position of the context circles, and their varying proximities to the target circle. This proposal is similar to the one used previously to explain the results of studies observing differences in grip aperture when grasping; participants may treat pictorially presented context circles as obstacles or distractors when grasping a central disk or "chip" and respond by adjusting their grip apertures ([Haffenden and Goodale, 2000](#); [Haffenden et al., 2001](#)). In this study, the context circles comprising the Perceived Large target may have been more likely to be treated as "obstacles" or "distractors" due to their close proximity to the target, in comparison to the farther positioned context circles of the Perceived Small target. If participants were treating the context circles as obstacles, they may have generated more directional changes when clicking the Perceived Large target to avoid a "collision" between the

cursor and context circles. It seems unlikely, however, that the context circles in the current experiment would elicit such obstacle avoidance mechanisms, as there was no associated risk of collision, and participants were not given any instruction to avoid the context circles. An alternative possibility is that the proximity of the context circles surrounding the Perceived Large target may have provided participants with a larger "general area" (i.e., combining the target and context circles) to which initial cursor movements were directed to. Additional directional changes would therefore be required during later stages of the movement, once the center target is considered separate from the entire configuration. Unfortunately, removal of the Perceived Large (Far) target from the analysis meant we could not distinguish the illusory effect induced by the context circles and their proximity to the target.

Participants demonstrated shorter deceleration phases when clicking on the Control Large target in comparison to the Perceived Small target in Experiment 2, and a significant effect of Target Type in Experiment 3 suggested deceleration phases were shortest in response to the Control Large target (however, Experiment 3 pairwise comparisons did not reach statistical significance). Observing shorter deceleration phases in response to the veridically larger Control Large target makes intuitive sense, as the larger target requires less precision at later stages of the movement. These results support other research demonstrating shorter deceleration phases toward targets perceived as larger ([Handlovsky et al., 2004](#)). However, it is less clear why the Perceived Small target in particular would generate longer deceleration phases in comparison. While it would be reasonable to expect longer deceleration phases for targets perceived as smaller, and therefore requiring increased precision, this was not the case for the veridically smaller Control Small target. If cursor movements directed toward the Perceived Small target benefited from a more accurate planning phase as suggested above, then in fact we might expect shorter deceleration phases when clicking on this target type, rather than the longer deceleration phases observed in Experiment 2. It should be noted, however, that cursor movements such as those generated in this experiment have the potential to be much more sporadic and unorganized than the types of reaching-to-grasp or pointing movements typically investigated (see [Figure 3](#) for an example of the variability of movement trajectories between trials). As a result, the deceleration phase (as defined in this study as the proportion of the movement following peak velocity) of a cursor movement may not be as reliable an indicator of online control or an increase in precision as it is when considering visually guided hand movements.

The results of this experiment also provide valuable information about how the target's onscreen position influenced participants' cursor movement. In general, participants' cursor movements were slower, more curved, and consisted of a higher number of corrective movements when the target was presented on the right side of the screen than compared to when presented

on the left. The time at which the maximum deviation from the idealized trajectory path occurred differed between left and right targets in the first block of trials in Experiment 2. Following the manipulation, however (increased emphasis on the speed at which the task was performed), this difference disappeared. The fact that these effects were observed regardless of the type of target being presented suggest these changes in the speed and trajectory of the cursor movements toward leftward and rightward presented targets are more likely a result of all participants controlling the cursor with their right hand, rather than the result of the illusory context.

Whereas, ipsilateral reaching movements are typically faster and more accurate than contralateral movements (Carson et al., 1993; Hodges et al., 1997) the additional mechanical constraints present when using the right hand to perform a rightward cursor movement on a trackpad introduces certain difficulties which may have contributed to the effects observed in this study. For example, using the right hand to perform a leftward movement of the cursor simply requires the extension of the digit on the touchpad, while a rightward movement requires adduction of the index finger (or rightward abduction of the middle finger), as well as a necessary adduction of the wrist. These results suggest future investigations of trackpad-controlled cursor movement and human-computer interaction in general should consider the added mechanical constraints associated with a rightward compared to a leftward cursor movement.

Methodological considerations

There are several methodological considerations that may have also contributed to the absence of an observed effect of the illusion on the performance variables in this study. First, the task itself was relatively easy, simply requiring participants to move their cursors toward the target and click the center quickly and accurately. Despite encouraging participants to perform the task faster in Experiment 2 and more accurately in Experiment 3, the actual task requirements did not effectively change in these experiments. Additionally, while the type of target varied between each trial, the target only ever appeared at one of two onscreen positions, on either the left or right side. This meant that regardless of the type of target presented, the center of the target was always located at the same leftward or rightward position. Therefore, even if the perceived size of the target did affect participants' performance of the task, the simple, repetitive nature of the task and participants' overall high performance may have masked the influence of the illusion.

Second, participants performed the task in a closed-loop fashion and visual feedback of the target was always available. Participants therefore had ample opportunity to refine and adjust their movements online to achieve a consistently high level of accuracy. Previous research successfully demonstrating an influence of the Ebbinghaus illusion on visually guided

aiming movements has typically involved removal of visual feedback to some degree, either by removing vision of the hand (van Donkelaar, 1999), or the target (Fischer, 2001; Alphonsa et al., 2016). Both the perception-action and planning-control models predict that by removing visual feedback of the target, a greater emphasis will be placed on participants' sensorimotor memory of the target's position, therefore recruiting the perceptual system's involvement in the task, and increasing the likelihood of an illusory influence. According to the perception-action model, the action toward the target's remembered location will rely primarily on stored representations of the target within the perceptually dominated ventral stream and will therefore be more susceptible to the original illusory context of the target. Similarly, the planning-control model suggests that without visual feedback of the target facilitating the online corrections occurring during the action's "control" phase, the movement will be primarily guided by the representation of the target's position constructed during the "planning" phase, which is also susceptible to the illusory context prior to effector movement.

Third, the perceptual comparison task used in this study to confirm if the illusions effectively induced a change in perceived target size may have influenced the strength of the illusion differently for the perceptual task compared to the movement task. For example, this study utilized a forced choice task between two target stimuli. It may be argued that the division of attention required for this type of task may be more likely to produce an illusory effect (Pavani et al., 1999; Franz et al., 2000; Foster and Franz, 2014), while the directed focus during an action toward a single target may reduce the influence of the illusion. It is therefore a possibility that the illusory stimuli used in this study were more effective during the perceptual comparison task than during the experimental trials, which could also explain the lack of any observed influence of the illusion when acting on the single target. Conversely, as participants were not constrained regarding the amount of time required to perform the perceptual comparisons, it is possible the potential for increased inspection time may have in fact weakened the effect of the illusion for those who spent more time viewing the targets (Bressan and Kramer, 2021).

Finally, the online nature of this experiment involved participants performing the experimental task remotely, using a wide range of device types, screen sizes and resolutions, and without experimenter supervision. As such, the presentation of the stimuli likely varied to some degree depending on the screen size and resolution of the device used by participants to complete the experiment. For example, the container in which the onscreen display was presented was set as relatively small (800 × 600 px) to accommodate the presentation of the experiment on a wide variety of screen sizes. The strength of the Ebbinghaus illusion has shown to increase as the size of the stimuli increase (Massaro and Anderson, 1971;

Knol et al., 2015) and therefore the smaller display may have weakened the influence of the illusion. Each experiment was conducted using a within-subjects repeated measures design as an attempt to control for the variety of display presentations between participants. However, the minimized experimental control inherent in this form of remote data collection remains a possible threat to the internal validity of the experiment.

Conclusion

Here we have provided an example of how remote data collection methods may be used to conduct behavioral research outside of the laboratory setting. Using the Ebbinghaus illusion, participants' perceptions of the onscreen stimuli were influenced during a point-and-click task performed remotely using their own laptop devices. While the trajectory of the cursor movement appeared to be influenced by the perceived size of the target illusory context, end-point measures such as click-point accuracy and movement time were not influenced by the illusory context. Despite the convenience provided by remote data collection, those utilizing remote data collection should carefully consider the appropriate methodological considerations and anticipate the potential decrease in experimental control associated with conducting behavioral research outside of the laboratory environment.

Data availability statement

The original contributions presented in this study are publicly available. This data can be found here: [doi: 10.34990/FK2/NC0MPM](https://doi.org/10.34990/FK2/NC0MPM).

Ethics statement

The studies involving human participants were reviewed and approved by the Psychology/Sociology Research Ethics Board (P/SREB) at the University of Manitoba. Written informed consent to participate in this study was provided by the participants.

References

- Aglioti, S., DeSouza, J. F. X., and Goodale, M. A. (1995). Size-contrast illusions deceive the eye but not the hand. *Curr. Biol.* 5, 679–685. doi: 10.1016/S0960-9822(95)00133-3
- Alphonsa, S., Dai, B., Benham-Deal, T., and Zhu, Q. (2016). Combined visual illusion effects on the perceived index of difficulty and movement outcomes

Author contributions

RL was responsible for construction of the online experiment, as well as collection and analysis of the data, and the writing of the original manuscript draft. JM provided funding acquisition, project resources, and supervision. Both authors contributed to the conception and design of the study, as well as revision of the submitted manuscript.

Funding

This research was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (RGPIN 04964-18) held by JM.

Acknowledgments

We thank Bradley Smith of the University of Manitoba for his assistance constructing and hosting the online experiment.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

in discrete and continuous Fitts' tapping. *Psychol. Res.* 80, 55–68. doi: 10.1007/s00426-014-0641-x

Borchers, H. W. (2021). *Pracma: Practical Numerical Math Functions*. R Package Version 2.3.3. Available Online at: <https://cran.r-project.org/package=pracma>

- Bressan, P., and Kramer, P. (2021). Most findings obtained with untimed visual illusions are confounded. *Psychol. Sci.* 32, 1238–1246. doi: 10.1177/0956797621994268
- Carson, R. G., Goodman, D., Chua, R., and Elliot, D. (1993). Asymmetries in the regulation of visually guided aiming. *J. Mot. Behav.* 25, 21–32. doi: 10.1080/00222895.1993.9941636
- Danckert, J. A., Sharif, N., Haffenden, A. M., Schiff, K. C., and Goodale, M. A. (2002). A temporal analysis of grasping in the Ebbinghaus illusion: Planning versus online control. *Exp. Brain Res.* 144, 275–280. doi: 10.1007/s00221-002-1073-1
- De Grave, D. D. J., Biegstraaten, M., Smeets, J. B. J., and Brenner, E. (2005). Effects of the Ebbinghaus figure on grasping are not only due to misjudged size. *Exp. Brain Res.* 163, 58–64. doi: 10.1007/s00221-004-2138-0
- Firebase. (2021). *Make Your App The Best It Can Be*. Available Online at: <https://firebase.google.com/>
- Fischer, M. H. (2001). How sensitive is hand transport to illusory context effects? *Exp. Brain Res.* 136, 224–230. doi: 10.1007/s002210000571
- Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *J. Exp. Psychol.* 47, 381–391.
- Foster, R. M., and Franz, V. H. (2014). Superadditivity of the Ebbinghaus and Müller-Lyer illusions depends on the method of comparison used. *Perception* 43, 783–795. doi: 10.1068/p7802
- Franz, V. H. (2003). Planning versus online control: Dynamic illusion effects in grasping? *Spat. Vis.* 16, 211–223. doi: 10.1163/15685680322467491
- Franz, V. H., Bühlhoff, H. H., and Fahle, M. (2003). Grasp effects of the Ebbinghaus illusion: Obstacle avoidance is not the explanation. *Exp. Brain Res.* 149, 470–477. doi: 10.1007/s00221-002-1364-6
- Franz, V. H., and Gegenfurtner, K. R. (2008). Grasping visual illusions: Consistent data and no dissociation. *Cogn. Neuropsychol.* 25, 920–950. doi: 10.1080/02643290701862449
- Franz, V. H., Gegenfurtner, K. R., Bühlhoff, H. H., and Fahle, M. (2000). Grasping visual illusions: No evidence for a dissociation between perception and action. *Psychol. Sci.* 11, 20–25. doi: 10.1111/1467-9280.00209
- Franz, V. H., Scharnowski, F., and Gegenfurtner, K. R. (2005). Illusion effects on grasping are temporally constant not dynamic. *J. Exp. Psychol.* 31, 1359–1378. doi: 10.1037/0096-1523.31.6.1359
- Freud, E., Macdonald, S. N., Chen, J., Quinlan, D. J., Goodale, M. A., and Culham, J. C. (2018). Getting a grip on reality: Grasping movements directed to real objects and images rely on dissociable neural representations. *Cortex* 98, 34–48. doi: 10.1016/j.cortex.2017.02.020
- Gilster, R., Kuitz-Buschbeck, J. P., Wiesner, C. D., and Ferstl, R. (2006). Grasp effects of the Ebbinghaus illusion are ambiguous. *Exp. Brain Res.* 171, 416–420. doi: 10.1007/s00221-006-0463-1
- Github. (2021). *Let's Build From Here, Openly Instantly Automatically Securely Magically Collaboratively Together*. Available Online at: <https://github.com/>
- Glover, S. (2004). Separate visual representations in the planning and control of action. *Behav. Brain Sci.* 27, 3–24. doi: 10.1017/s0140525x04000020
- Glover, S., and Dixon, P. (2002). Dynamic effects of the Ebbinghaus illusion in grasping: Support for a planning/control model of action. *Percept. Psychophys.* 64, 266–278. doi: 10.3758/BF03195791
- Goodale, M. A., and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25. doi: 10.1016/0166-2236(92)90344-8
- Goodale, M. A., and Milner, A. D. (2004). Plans for Action. *Behav. Brain Sci.* 27, 3–24.
- Haffenden, A. M., and Goodale, M. A. (1998). The effect of pictorial illusion on perception and visually guided prehension. *Investig. Ophthalmol. Vis. Sci.* 37, 122–136. doi: 10.1162/089892998563824
- Haffenden, A. M., and Goodale, M. A. (2000). Independent effects of pictorial displays on perception and action. *Vis. Res.* 40, 1597–1607. doi: 10.1016/S0042-6989(00)00056-0
- Haffenden, A. M., Schiff, K. C., and Goodale, M. A. (2001). The dissociation between perception and action in the Ebbinghaus illusion: Nonillusory effects of pictorial cues on grasp. *Curr. Biol.* 11, 177–181. doi: 10.1016/S0960-9822(01)00023-9
- Handlovsky, I., Hansen, S., Lee, T. D., and Elliott, D. (2004). The Ebbinghaus illusion affects on-line movement control. *Neurosci. Lett.* 366, 308–311. doi: 10.1016/j.neulet.2004.05.056
- Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., and Hilbig, B. E. (2021). lab.js: A free, open, online study builder. *Behav. Res. Methods* 54, 556–573. doi: 10.3758/s13428-019-01283-5
- Hodges, N. J., Lyons, J., Cockell, D., Reed, A., and Elliot, D. (1997). Hand, space and attentional asymmetries in goal-directed manual aiming. *Cortex* 33, 251–269. doi: 10.1016/s0010-9452(08)70003-0
- Janczyk, M., Pfister, R., and Kunde, W. (2013). Mice move smoothly: Irrelevant object variation affects perception, but not computer mouse actions. *Exp. Brain Res.* 231, 97–106. doi: 10.1007/s00221-013-071-5
- Johnson, A., and Baddeley, A. (2019). *Polycclip: Polygon Clipping, R Package Version 1.10-0*. Available Online at: <https://cran.r-project.org/package=polycclip>
- Kieslich, P. J., and Henninger, F. (2017). Mousetrap: An integrated, open-source mouse-tracking package. *Behav. Res. Methods* 49, 1652–1667. doi: 10.3758/s13428-017-0900-z
- Kieslich, P. J., Henninger, F., Wulff, D. U., Haslbeck, J. M. B., and Schulte-Mecklenbeck, M. (2019). “Mouse-tracking: A practical guide to implementation and analysis,” in *A Handbook of Process Tracing Methods: Second Edition*, eds M. Schulte-Mecklenbeck, A. Kühberger, and J. G. Johnson (New York, NY: Routledge), 111–130. doi: 10.31234/osf.io/zuvqa
- Knol, H., Huys, R., Sarrazin, J., and Jirsa, V. K. (2015). Quantifying the Ebbinghaus figure effect: Target size, context size, and target-context distance determine the presence and direction of the illusion. *Front. Psychol.* 6:1679. doi: 10.3389/fpsyg.2015.01679
- Knol, H., Huys, R., Sarrazin, J., Spiegler, A., and Jirsa, V. K. (2017). Ebbinghaus figures that deceive the eye do not necessarily deceive the hand. *Sci. Rep.* 7:3111. doi: 10.1038/s41598-017-02925-4
- Kopiske, K. K., Bruno, N., Hesse, C., Schenk, T., and Franz, V. H. (2016). The functional subdivision of the visual brain: Is there a real illusion effect on action? A multi-lab replication study. *Cortex* 79, 130–152. doi: 10.1016/j.cortex.2016.03.020
- Li, Q., Joo, S. J., Yeatman, J. D., and Reinecke, K. (2020). Controlling for participants' viewing distance in large-scale psychophysical online experiments using a virtual chinrest. *Sci. Rep.* 10:904. doi: 10.1038/s41598-019-57204-1
- Marotta, J. J., DeSouza, J. F. X., Haffenden, A. M., and Goodale, M. A. (1998). Does a monocularly presented size-contrast illusion influence grip aperture? *Investig. Ophthalmol. Vis. Sci.* 37, 380–386. doi: 10.1016/s0028-3932(97)00154-1
- Massaro, D. W., and Anderson, N. H. (1971). Judgmental model of the Ebbinghaus illusion. *J. Exp. Psychol.* 89, 147–151.
- Milner, A. D., and Goodale, M. A. (2006). *The Visual Brain in Action*, 2nd Edn. Oxford: Oxford University Press.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4
- Ozana, A., Namdar, G., and Ganel, T. (2020). Active visuomotor interactions with virtual objects on touchscreens adhere to Weber's law. *Psychol. Res.* 84, 2144–2156. doi: 10.1007/s00426-019-01210-5
- Pavani, F., Boscagli, I., Benvenuti, F., Rabuffetti, M., and Farnè, A. (1999). Are perception and action affected differently by the Titchener circles illusion? *Exp. Brain Res.* 127, 95–101. doi: 10.1007/s002210050777
- Phillips, J. G., Triggs, T. J., and Meehan, J. W. (2003). Conflicting directional and locational cues afforded by arrowhead cursors in graphical user interfaces. *J. Exp. Psychol.* 9, 75–87. doi: 10.1037/1076-898X.9.2.75
- Phillips, J. G., Triggs, T. T., and Meehan, J. W. (2001). Cursor orientation and computer screen positioning movements. *Hum. Factors* 43, 435–441. doi: 10.1518/001872001775898241
- R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R foundation for Statistical Computing.
- Roberts, B., Harris, M. G., and Yates, T. A. (2005). The roles of inducer size and distance in the Ebbinghaus illusion (Titchener circles). *Perception* 34, 847–857.
- Sutter, C., Müseler, J., and Bardos, L. (2011). Effects of sensorimotor transformations with graphical input devices. *Behav. Inf. Technol.* 30, 415–424. doi: 10.1080/01449291003660349
- van Donkelaar, P. (1999). Pointing movements are affected by size-contrast illusions. *Exp. Brain Res.* 125, 517–520. doi: 10.1007/s002210050710
- Westwood, D. (2004). Planning, control, and the illusion of explanation. *Behav. Brain Sci.* 27, 54–55. doi: 10.1017/S0140525X04490025



OPEN ACCESS

EDITED AND REVIEWED BY
Jan M. Hondzinski,
Louisiana State University,
United States

*CORRESPONDENCE
Ryan W. Langridge
langrirw@myumanitoba.ca

SPECIALTY SECTION
This article was submitted to
Perception Science,
a section of the journal
Frontiers in Psychology

RECEIVED 13 September 2022
ACCEPTED 25 October 2022
PUBLISHED 15 November 2022

CITATION
Langridge RW and Marotta JJ (2022)
Corrigendum: Use of remote data
collection methodology to test for an
illusory effect on visually guided cursor
movements.
Front. Psychol. 13:1042774.
doi: 10.3389/fpsyg.2022.1042774

COPYRIGHT
© 2022 Langridge and Marotta. This is
an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction
in other forums is permitted, provided
the original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Corrigendum: Use of remote data collection methodology to test for an illusory effect on visually guided cursor movements

Ryan W. Langridge* and Jonathan J. Marotta

Perception and Action Lab, Department of Psychology, University of Manitoba, Winnipeg, MB, Canada

KEYWORDS

Ebbinghaus illusion, Titchener circles, cursor control, perception, action

A corrigendum on

Use of remote data collection methodology to test for an illusory effect on visually guided cursor movements

by Langridge, R. W., and Marotta, J. J. (2022). *Front. Psychol.* 13:922381.
doi: 10.3389/fpsyg.2022.922381

In the published article, there was an error in [Figure 4](#) as published. The Y-Axis Label incorrectly stated: Number of Corrective Movements. The correct Y-Axis Label is: Number of Directional Changes. The corrected [Figure 4](#) and its caption appear below.

In the published article, there was an error in [Figure 7](#) as published. The Y-Axis Label incorrectly stated: Number of Corrective Movements. The correct Y-Axis Label is: Number of Directional Changes. The corrected [Figure 7](#) and its caption appear below.

In the published article, there was an error in [Figure 10](#) as published. The Y-Axis Label incorrectly stated: Number of Corrective Movements. The correct Y-Axis Label is: Number of Directional Changes. The corrected [Figure 10](#) and its caption appear below.

The authors apologize for these errors and state that they do not change the scientific conclusions of the article in any way. The original article has been updated.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

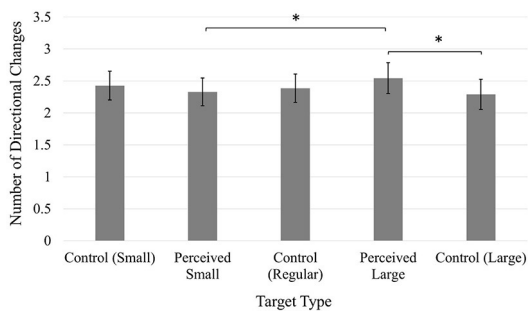


FIGURE 4
Average number of directional changes. Error bars represent 95% confidence intervals. $p < 0.05$.

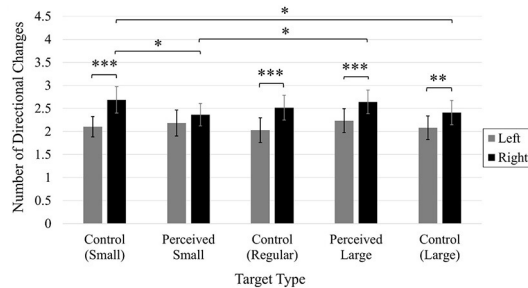


FIGURE 7
Average number of directional changes. Error bars represent 95% confidence intervals. $p < 0.05$, $**p < 0.01$, $***p < 0.001$.

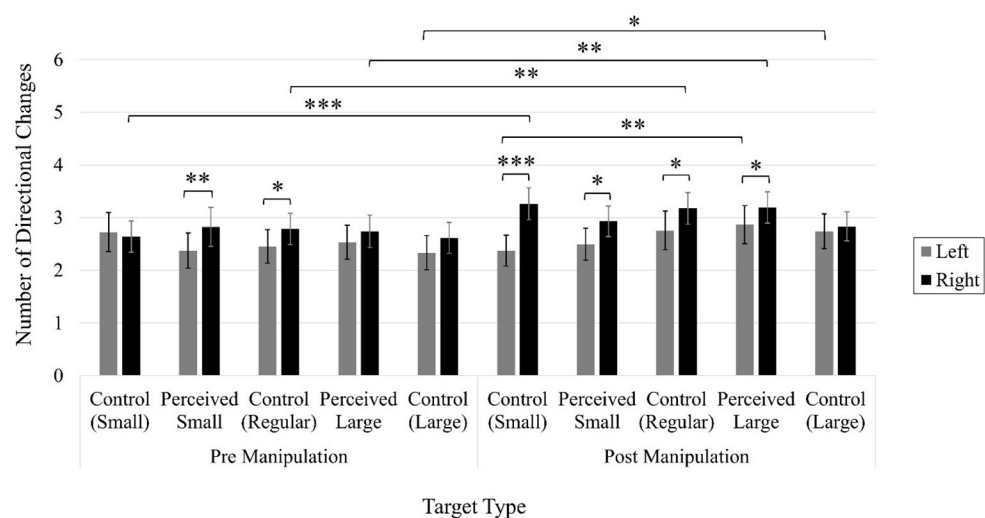


FIGURE 10
Average number of directional changes. Error bars represent 95% confidence intervals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.



OPEN ACCESS

EDITED BY

Alyssa A. Brewer,
University of California, Irvine,
United States

REVIEWED BY

Ming Luo,
Zhejiang University, China
Peter Rhodes,
University of Leeds, United Kingdom

*CORRESPONDENCE

Yiting Duan
yitingduan@163.com
Min Zhang
zhangmin306@126.com

SPECIALTY SECTION

This article was submitted to
Perception Science,
a section of the journal
Frontiers in Psychology

RECEIVED 30 May 2022

ACCEPTED 29 August 2022

PUBLISHED 18 October 2022

CITATION

Duan Y, Zhang M, Zhou J, Zhou H, Xu L
and Yang H (2022) Exploring the law of
color presentation of double-sided
heterochromatic digital printing for
textile. *Front. Psychol.* 13:956748.
doi: 10.3389/fpsyg.2022.956748

COPYRIGHT

© 2022 Duan, Zhang, Zhou, Zhou, Xu
and Yang. This is an open-access
article distributed under the terms of
the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution
or reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Exploring the law of color presentation of double-sided heterochromatic digital printing for textile

Yiting Duan^{1,2,3*}, Min Zhang^{1*}, Jiu Zhou¹, Hua Zhou²,
Liming Xu² and Heng Yang¹

¹College of Textile Science and Engineering, International Silk Institute, Zhejiang Sci-Tech University, Hangzhou, China, ²Hangzhou Honghua Digital Technology Co., Ltd., Hangzhou, China, ³School of Design, University of Leeds, Leeds, United Kingdom

With the upgrading of production technology and design aesthetic of textile products, the development and application of double-sided digital printing technology have been gradually proposed in recent years. The production of double-sided heterochromatic digital printing requires high accuracy for color management, as in the production process, different input devices, display devices, and output devices each have varying color processing capabilities and color performance characteristics, which leads to the transfer of color between different devices and not being accurately reproduced. However, there is little research involved in the color formation laws in the design and production of double-sided heterochromatic digital printing. The purpose of this paper is to explore the prediction model of color presentation in double-sided heterochromatic digital printing. Due to the influence of the fabric thickness, the gap between the textile structure, and the infiltration rate of the printing pigments, the color of each side in double-sided, heterochromatic printing results often differs from the designed color. In this paper, taking chiffon fabric, which is one of the thinnest and has the strongest permeability in silk fabric, as an example fabric base, 24 colors from six hue angles and four chromas were selected as experimental colors. According to the color gamut of the digital printing machine, 15 out of 24 colors were selected as experimental colors. These 15 colors were printed on two sides in pairs to generate 225 color pairs as double-sided experimental samples. For these experimental samples, this paper conducts experiments from both subjective and objective aspects through the combination of subjective evaluation of psychophysics experiments and objective instrument measurement. Through the analysis of experimental data, the color difference prediction under the subjective model ($R^2 = 0.74$) and objective model ($R^2 = 0.85$) are given, respectively. The dominant color prediction model ($R^2 = 0.75$) during double-sided heterochromatic digital printing has also been

built. The combination of the three models can predict the color regularity of double-sided heterochromatic digital printing on silk, which may have a certain significance for the design and development of silk double-sided heterochromatic digital printing products.

KEYWORDS

double-sided heterochromatic digital printing, textile digital printing, subjective color difference, objective color difference, regression model

Introduction

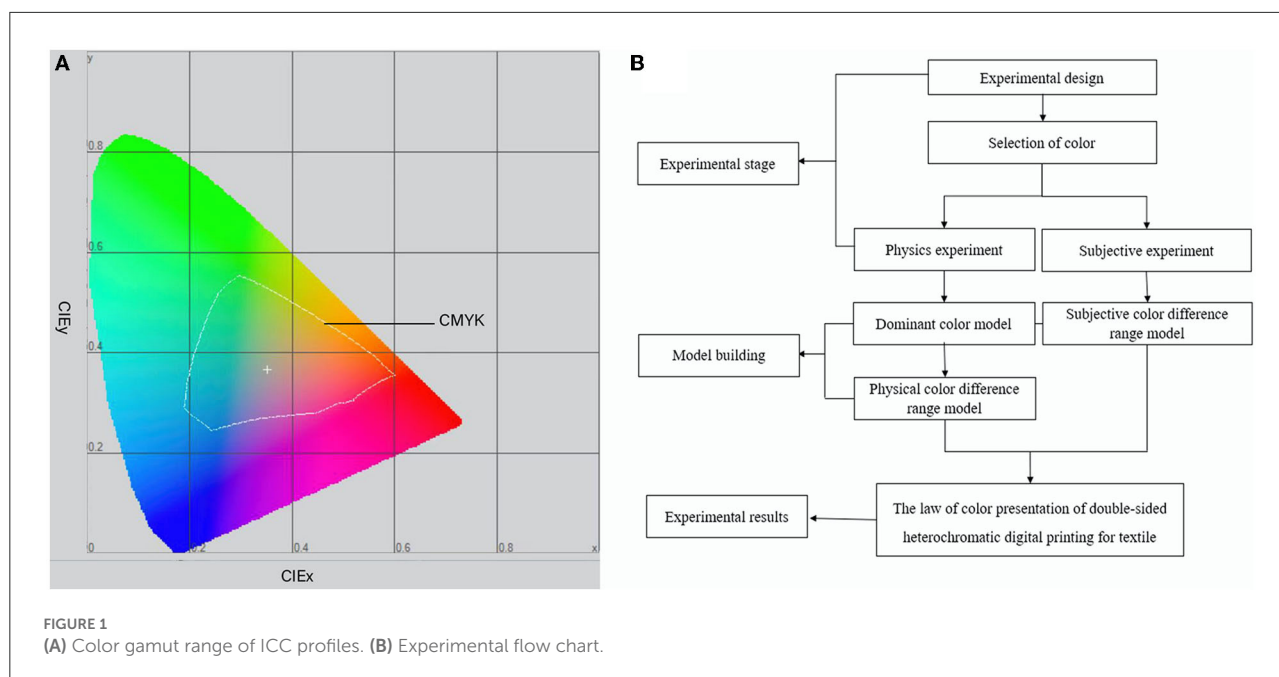
As people's living standard improves, customers' assessment of textiles is expanding to include not only the texture of the textile material, but also the need for its color. Color is a vital component in human visual information transmission, since it not only delivers the designer's message but also reflects people's feelings and has a positive influence on their spirit and conduct (Ou et al., 2012). The color of textile patterns has become one of the most essential variables for customers to consider when buying textiles, particularly among young consumers who are fashion savvy and individualistic. According to the research, more than 40% of customers' first consideration when purchasing textiles is color. If the preferred textiles do not have their favorite color, they will opt to temporarily not buy or buy others instead, as it is usually not easy for people to give up their favorite colors (Wang et al., 2016). In American marketing, there is a 7-s rule that states that customers will determine their buying intention within 7 s, and color subjective accounts for 67% of the decision element in such a short period (Qin, 2014). This demonstrates the significance of color in consumer behavior, and the influence of textile color in affecting product look is even more intangible.

To fulfill the need for manufacturing and design, textile digital printing technology, a complete high technology combining computer, electronic information, machinery, and other disciplines, was developed in the 1990's (Yang et al., 2010). Digital printing technology is a significant achievement in the world of textile printing, and it is the result of societal development and current trends. Digital printing is the result of the organic combination of printing technology and computer technology, and is based on computer graphics, digital manufacturing, and computer networks. Image information is input into the computer, edited, and calculated by the color separation system, and then the dye is directly sprayed onto the fabric or media by the printing system to produce beautiful printing patterns (Tyler, 2005). On the one hand, the problem with the current situation of digital printing color management in China is that the technology of digital printing color management is primarily imitated from the paper printing industry, whereas digital printing of textiles is very dissimilar from paper printing, whether it is hardware equipment, ink,

fabric, or process, and textile digital printing is much more complex than paper printing (Gooby, 2020). As a result, the advancement of color management technology in the printing sector needs to be paid attention, but it cannot be entirely replicated. Rather, the peculiarities of digital textile printing need to be enhanced and expanded. Another issue is that color difference is at the heart of textile quality, and the role of color difference measurement in the process of textile quality evaluation is becoming abundantly obvious (Gangakhedkar, 2010), but there is a scarcity of literature research and systematic application of double-sided printing color difference detection about silk fabric. There has been limited research, in particular, on color location and color creation criteria in color design and production in double-sided heterochromatic digital printing. Because of the thinness of the fabric, gaps between the tissue structure, and the permeability of the printing pigments, the printed color in double-sided digital printing often differs from the planned hue angle (Tyler, 2011). Color differences occur in all color reproduction sectors, and this phenomenon is most noticeable in chiffon textiles, thus the purpose of this research is to explore the color pattern of double-sided heterochromatic digital printing with high permeability. The experimental fabric used in this research is a high-permeability chiffon fabric. This research uses a combination of the subjective evaluation method of psychophysics experiment and instrumental measurement method to conduct the experiment from both subjective and objective perspectives to look at the problem of color difference in chiffon fabric when doing double-sided heterochromatic digital printing. The three models combined can forecast the color presentation pattern of double-sided heterochromatic digital printing on silk, which may be useful in the design and development of silk double-sided heterochromatic digital printing goods.

Materials and methods

This paper primarily uses a combination of objective measurement and subjective measurement to analyze the color presentation pattern of chiffon fabric when doing double-sided heterochromatic digital printing, and it provides a prediction model of color difference size based on subjective and objective measurements. In order to make the color



printed by the digital printer to be closer to the designed color, the design colors were color managed by using the ICC profile of the digital printing machine provided by Hangzhou Honghua Digital Technology Co., Ltd., and the color gamut range of this ICC is shown in Figure 1A. A total of 210 colors from the digital printing machine's color range were chosen as experimental samples, and objective measurement and subjective measurement studies were performed on these experimental samples. The findings of 30 individuals assessing the degree of color difference between 210 single- and double-sided experimental samples under the visual influence solely were gathered in the subjective experiment. In the objective experiments, a CM700d spectrophotometer was used to measure the $L^*a^*b^*$ chromaticity values of the front and back sides of the experimental samples, and the color difference values (hereinafter referred to as ΔE^*) of the corresponding colors of the single- and double-sided experimental samples are calculated using the CIEDE2000 color difference formula based on the CIELAB color space. Finally, this paper analyzes the causes for the color difference of the experimental samples in subjective and objective assessments, as well as the connection of color difference size in subjective and objective evaluations. In this method, the color presentation prediction model during double-sided heterochromatic digital printing of textiles is determined. Figure 1B depicts the experimental flow.

The choice of color

In this paper, 24 colors are chosen from six hue angles and four color chroma, and the colors are distributed in CIELAB

color space with the equal lightness of a^*-b^* two-dimensional distribution, as shown in Figure 2A, in one, two, and four quadrants, each with eight colors, and the 24 colors are grouped in pairs (two colors in each group) to generate a total of 576 different colors. In order to ensure the authenticity and validity of the data, the specimen (single-sided chiffon color card) required for the experiment was folded in five layers according to the measurement method of textile color (Gangakhedkar, 2010), and the $L^*a^*b^*$ value of the specimen was measured by CM700d spectrophotometer (the specular component is SCI) in the case of the fabric sample without light transmission. It was discovered that there is a significant disparity between the chromaticity value and the intended color. Given that the color of digital printing is mixed by four CMYK colors, and that digital printing is used to control the output color by computer, color management was performed in accordance with the ICC profile of the digital printing machine. Some colors were discovered to be beyond the color gamut of the printer, resulting in severe color differences. After removing the super gamut colors, 15 colors remained, and their chromaticity values $L^*a^*b^*$ are shown in Supplementary Table 1, and their two-dimensional distribution in CIELAB color space with equal lightness a^*-b^* is displayed in Figure 2B. They are positioned in one, two, and four quadrants, with 7 colors in the first quadrant and 4 colors in each of the second and fourth quadrants, and these 15 colors are paired in groups to produce 225 distinct color samples; among them, 210 colors are heterochromatic. Figure 2C depicts the a^*-b^* two-dimensional distribution of these 210 color samples with equal lightness in CIELAB color space, with $|a^*|$ and $|b^*|$ values in the (0, 70) interval; in the fourth quadrant, the $|a^*|$ values are in the (0, 30) interval, and the $|b^*|$ values are in the (0, 40)

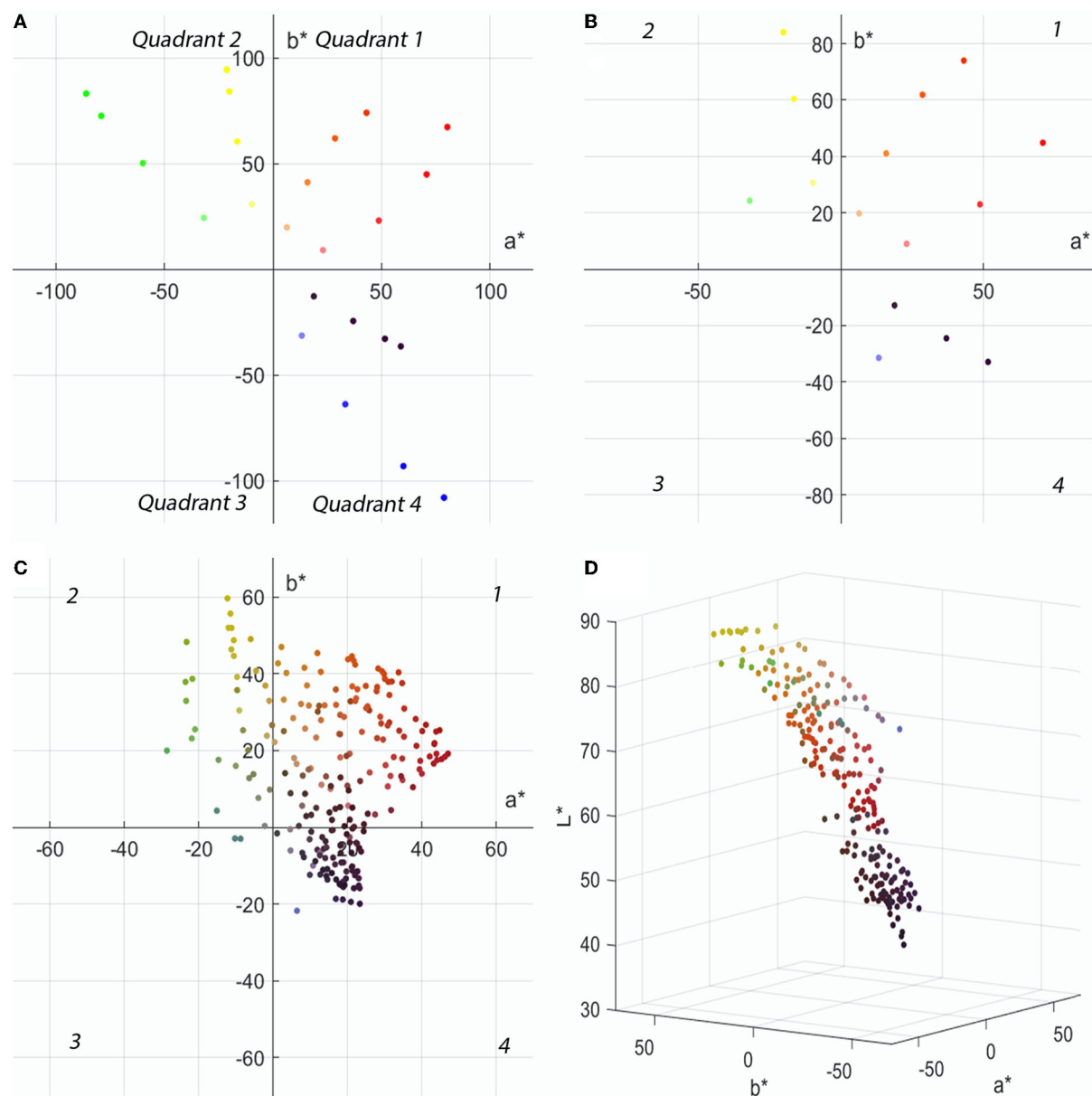


FIGURE 2

The color distribution in CIELAB space: (A) Two-dimensional distribution of 24 colors in CIELAB space. (B) Two-dimensional distribution of 15 colors in CIELAB space. (C) Two-dimensional distribution of 210 colors in CIELAB space. (D) Three-dimensional distribution of 210 colors in CIELAB space.

interval, which is red-blue with lower saturation. As shown in Figure 2D, most of the yellow samples have the highest lightness value between 70 and 90; green has the second highest lightness value, which is between 60 and 80; red lightness values are between 30 and 70, which demonstrate a larger interval; the lightness values of blue and purple samples are relatively low, yet they are more than 30. Due to the limitations related to the experimental material and the duration of the experiment, only 210 color samples were evaluated in this paper.

Materials and methods

Psychophysics experimental method and instrument objective measurement method are primarily employed to subjectively and objectively analyze the color presentation pattern of chiffon fabrics in conducting double-sided heterochromatic digital printing, while providing a prediction model of the level of color difference under subjective and objective measurements.



The objective color assessment approach in this research is to use a CM700d spectrophotometer to evaluate the chromaticity values $L^*a^*b^*$ of 210 colors based on CIELAB color space, as shown in [Figure 3](#). The samples are measured in five layers to ensure that the samples are opaque; all samples were in a consistent environment and placed on the 90% reflective white plate for measurement. Before measuring, the spectrophotometer was set to D65 light source, d/8 (diffuse lighting 8° light receiving), specular inclusion (SCI), measuring aperture of 8 mm, and used a 10-degree CIE standard observer. In addition, the repeatability of Konica Minolta CM700d was $\Delta E^*ab \leq 0.04$ and the inter-instrument variation was $\Delta E^*ab \leq 0.2$, after which the white board and black board were calibrated, and the specimens were measured with the L^* , a^* , and b^* values of each color card recorded. The objective color difference (hereinafter referred to as ΔE^*) of the experimental samples was calculated using the standard color difference method CIEDE2000 in CIELAB space ([Luo et al., 2001](#); [Wang et al., 2012](#); [Westland and Pan, 2017](#)). The relationship between

The assessment of the quality of textile digital printing needs to be based on people's feelings, and the major and most accurate evaluation of this feeling is people's subjective evaluation. As a result, this subjective experiment estimates the extent of the color difference based on people's visual ratings. The experimental sample of double-sided heterochromatic digital printing was subjected to a visual experiment in order to investigate the relationship between the subjective perceptible color difference, color lightness, and color phase when doing double-sided heterochromatic digital printing on semi-transparent chiffon fabric. The sample size for the research was computed using G*power 3.1 software (Faul et al., 2007) with a medium effect size of 0.50, and the optimal sample size for the paired-sample *t*-test was 34 groups under statistical test power of $1-\beta = 0.80$ and $\alpha = 0.05$. This study includes 30

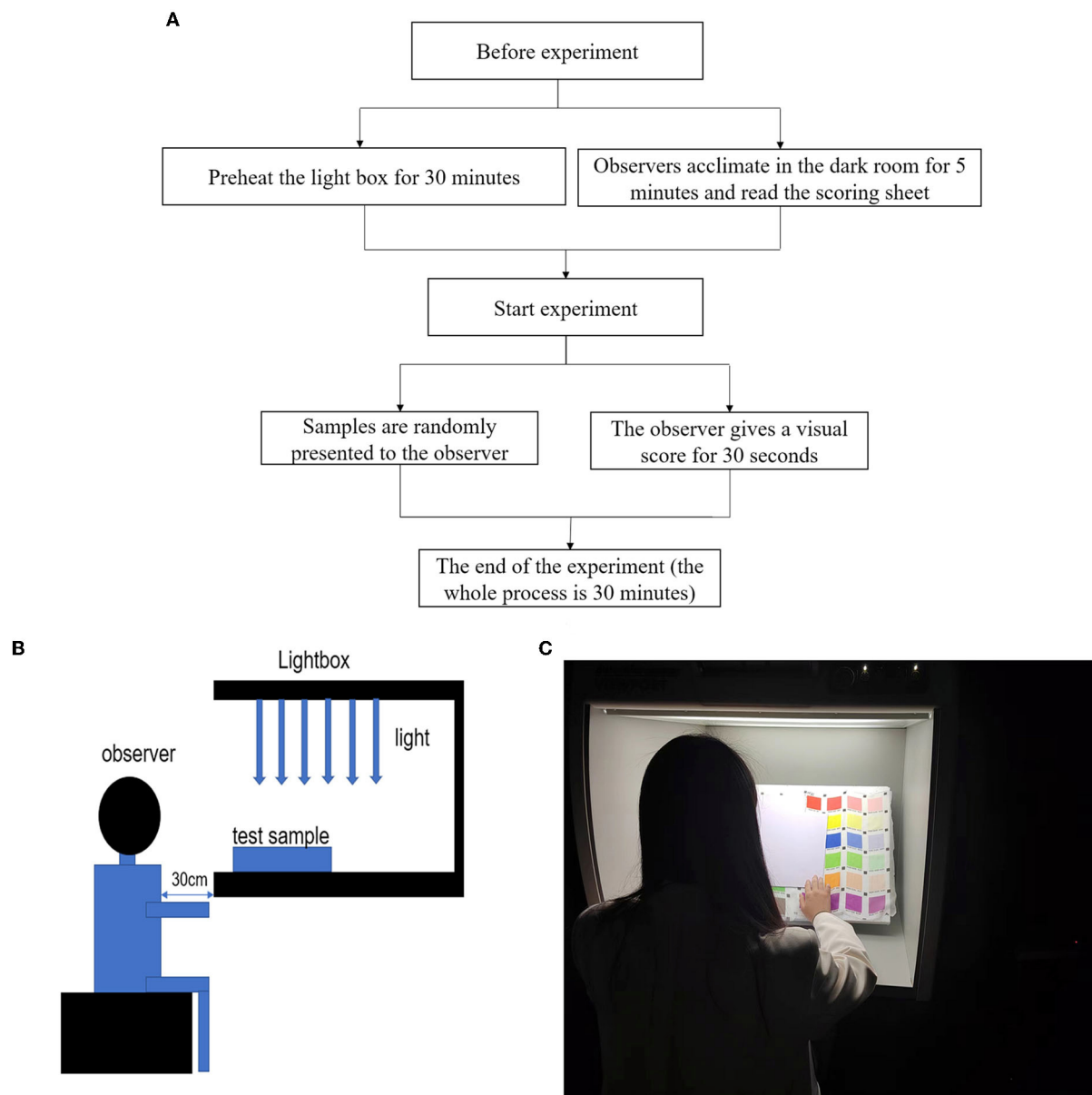


FIGURE 4
(A) Subjective experiment flow chart, (B) The experiment environment, and (C) Experiment procedure.

undergraduate and postgraduate students aged 20–25 years, all of whom have similar study background, and equal numbers of men and women were submitted to a repeat experiment after 1 week, and a total of 60 sets of experimental data were obtained. According to the Ishihara color blindness test, none of the 30 volunteers chosen for the experiment were color defiant. The experimental steps are shown in Figure 4A.

This test was conducted in a dark room with no other lighting source, the light source was D65, the illumination of the light box used in the experiment was 2250 under a D65

light source, and the light box was painted in standard gray (L^* of 50). The observer is required to maintain an upright sitting position while staring directly at the specimens in the light box, as indicated in Figures 4B,C. The observer (subjective evaluator) visually compares the 210 experimental samples of double-sided heterochromatic digital printing to the single-sided printed specimens using the scoring criteria, and assigns evaluation scores based on the magnitude of the scores to judge the size of the printed color difference. Ten color similarity score ranges were created to assess the double-sided heterochromatic

TABLE 1 Perception experiment scoring rules.

Grade	Score interval	Evaluation standard
1	[0.0–1.0]	There is almost no color difference
2	[1.1–3.0]	There is a slight color difference
3	[3.1–5.0]	There is obvious color difference
4	[5.1–7.0]	There is a very large color difference
5	[7.1–10.0]	Different color from standard

digital printed textiles. Table 1 shows the specifics of the scores.

At the same time, the test sample (double-sided heterochromatic digital printing) and standard sample (single-sided digital printing) color cards were put in a lightbox with a D65 light source (only one color was shown at a time, and when one color block was shown, the other color blocks were blocked with white cardboard). Before beginning the experiment, the light source D65 in the light box was switched on at full lightness for 30 min to warm up. Before entering the dark room, observers read the experimental procedures and scoring criteria and waited for 5 min to adjust to the dark environment before doing the visual experiment individually. The observer is required to tell the size of the color difference between the samples within 5 s, and the time between consecutive samples can be adjusted according to the observer's own state. The color difference of the samples in this experiment is completely subjectively determined by the observer and is not subject to external interference. For example, if the color between the sample and reference pairs was the same, the response would be "0"; if the color difference between the subject sample and the standard sample was different, the response would be a bigger number ranging from 0.1 to 10. The answer was recorded by the recorder and could not be corrected. One week later, the observers were asked to repeat the experiment. The specimens were given to the observer in random order for each observation to prevent the influence of the observer's memory effect on the experimental outcomes.

Results analysis of the objective color difference of double-sided heterochromatic digital printing

When double-sided digital printing was delivered, Pearson correlation analysis was performed between the color difference values and the lightness, chroma, and hue of the color, as well as the model of the dominant color and the model of the color difference range.

TABLE 2 Pearson correlation between ΔE^* and L^* , C_{ab}^* , and h_{ab} .

	L^*	C_{ab}^*	h_{ab}
Pearson correlation coefficient	−0.30**	−0.32**	0.07
Sig.	0.000	0.000	0.314

**Indicates that the test method used in this paper is a two-tailed significance test method.

Objective color difference law of double-sided heterochromatic digital printing

In this paper, the objective color difference ΔE^* (color difference between the single-sided control group and double-sided experiment) and $L^*C_{ab}^*h_{ab}$ were calculated using the CIEDE2000 formula in order to investigate the relationship between objective color difference and color lightness, chroma, and hue angle during double-sided heterochromatic digital printing, and the results of Pearson correlation analysis are presented in Table 2.

Pearson's correlation between ΔE^* , L^* , and C_{ab}^* was found to be significant, $\text{Sig.} = 0.000 \leq 0.001$. There was no substantial Pearson association between ΔE^* and h_{ab} . The Pearson correlation coefficients of ΔE^* with L^* and C_{ab}^* were −0.3 and −0.32, respectively, indicating a negative relationship between ΔE^* , lightness, and chroma, i.e., the smaller the lightness and chroma of the designed color within a certain range, the larger the resulting single- and double-sided color difference. To summarize, the objective color difference between single- and double-sided digital printing is largely impacted by lightness and chroma.

Determination of the dominant color during double-sided heterochromatic digital printing

Because chiffon is one of the thinnest silk textiles, the color presentation will be impacted by the color of the opposite side when undertaking double-sided heterochromatic digital printing. This paper examines the relationship between the color difference between the front and back sides of the double-sided experimental samples and the difference in $L^*C_{ab}^*h$ between the front and back sides to determine which colors dominate the presentation of fabric color when two different colors are matched to do double-sided heterochromatic digital printing. The difference of color difference between the front and back of the experimental sample is recorded as ΔE_{f-b}^* . When $\Delta E_{f-b}^* < 0$, it means that the objective color difference between the front side of the experimental sample and the designed color, ΔE_f^* , is smaller than the ΔE_b^* . That is, at this time, the color presentation is more biased toward the color of the

front of the experimental sample, and the front color is the dominant color. The difference in the L^*C^*h of the front and back specimens are recorded as ΔL^*_{f-b} , $\Delta C^*_{b(f-b)}$, and $\Delta h_{ab(f-b)}$. However, in the color system of L^*C^*h , when calculating the hue difference, it is generally expressed as ΔH^*_{ab} (Wang and Dou, 2005), and the Pearson correlation test was done between $\Delta L^*_{(f-b)}$, $\Delta C^*_{ab(f-b)}$, and $\Delta H^*_{ab(f-b)}$, as shown in Table 3. The results of Pearson correlation analysis showed that $\Delta E^*_{(f-b)}$ was positively correlated with $\Delta L^*_{(f-b)}$ and negatively correlated with $\Delta C^*_{ab(f-b)}$. That is, the larger $\Delta L^*_{(f-b)}$, the larger $\Delta E^*_{(f-b)}$, the larger the $\Delta C^*_{ab(f-b)}$, and the smaller the $\Delta E^*_{(f-b)}$.

Through the Pearson correlation, it is feasible to determine how to perform double-sided heterochromatic digital printing while the dominant color and the lightness difference between the front and back side are closely related, in order to investigate whether the dominant color presentation law can be expressed as ΔL^*_{f-b} , ΔC^*_{f-b} , and $\Delta H^*_{ab(f-b)}$. There is a linear relationship between ΔL^*_{f-b} , ΔC^*_{f-b} , and $\Delta H^*_{ab(f-b)}$, according to the regression analysis, and the regression equation can be represented as follows:

Dominant color determination

$$\Delta E^*_{(f-b)} = -0.102 + 4.554 \Delta L^*_{(f-b)} - 1.084 \Delta C^*_{ab(f-b)} + 0.024 \Delta H^*_{ab(f-b)} (R^2 = 0.745) \quad (1)$$

TABLE 3 Pearson correlation between ΔE^*_{f-b} and ΔL^*_{f-b} , ΔC^*_{f-b} , and $\Delta H^*_{ab(f-b)}$.

	ΔL^*_{f-b}	ΔC^*_{f-b}	$\Delta H^*_{ab(f-b)}$
Pearson correlation coefficient	0.80**	-0.60**	-0.01
Sig.	0.000	0.00	0.991

**Indicates that the test method used in this paper is a two-tailed significance test method.

Where ΔL^*_{d-s} is the lightness difference between the front and back, $\Delta C^*_{(d-s)}$ is the chroma difference between the front and back, and $\Delta H^*_{ab(f-b)}$ is the hue difference between the front and back. Significance analysis $R^2 = 0.75$ suggests that the model fits well and may be used to forecast the dominant color in double-sided heterochromatic digital printing. The results show that, when double-sided heterochromatic digital printing is applied to chiffon fabric, the color with the lowest lightness is the dominant color. When the lightness is fixed, the chroma influences the color pattern, and the color with the highest chroma being the dominant color. Furthermore, compared with lightness and chroma, the influence on hue is low according to determining the front and back color difference. As shown in Figure 5A, when the red and green (lightness of the green is higher than the red) are paired in groups to conduct double-sided heterochromatic digital printing, the color presented more in favor of the red. When the lower chroma green and higher chroma red are paired to generate two-sided heterochromatic digital printing, the color is more inclined to the higher chroma red, as illustrated in Figure 5B. When grouping two distinct hue angles, this model might assist the designer in predicting the dominating color.

The color difference range of double-sided heterochromatic digital printing

The previous section investigated the problem of determining the dominant color when doing double-sided

TABLE 4 Pearson correlation between ΔE^* and $\Delta L^*_{(d-s)}$, $\Delta C^*_{ab(d-s)}$, and $\Delta H^*_{ab(d-s)}$.

	$\Delta L^*_{(d-s)}$	$\Delta C^*_{ab(d-s)}$	$\Delta H^*_{ab(d-s)}$
Pearson correlation coefficient	-0.84**	-0.18**	0.75**
Sig.	0.000	0.000	0.000

**Indicates that the test method used in this paper is a two-tailed significance test method.

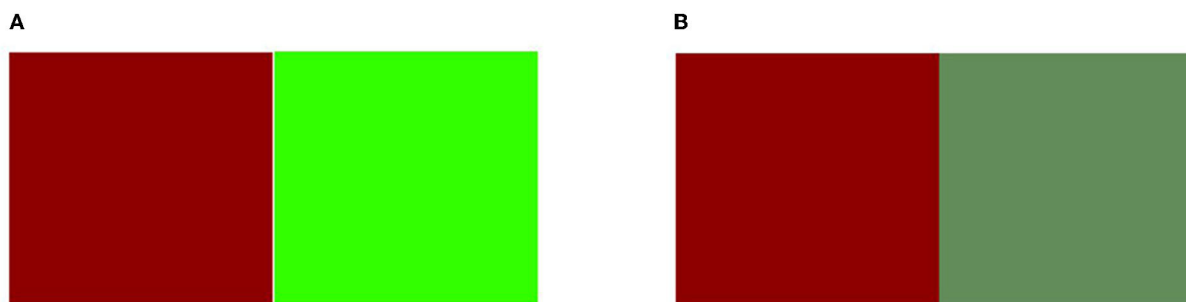


FIGURE 5
(A) Contrasting color pairs of the same chroma. (B) Contrasting color pairs of the same lightness.

heterochromatic digital printing in chiffon fabric. The following section investigates the color difference between double-sided heterochromatic digital printing and single-sided printing, thereby looking for whether there is some kind of pattern among the color difference rules. The color difference between the single- and double-sided experimental specimens is analyzed based on the difference between the $L_{ab}^* C_{ab}^* h_{ab}$ values of single- and double-sided colors. At the same time, the color difference between single- and double-sided experimental fabrics are calculated using the CIEDE2000 formula, is recorded as ΔE^* , and the difference between the $L_{ab}^* C_{ab}^* h_{ab}$ values of single- and double-sided specimens are recorded as $\Delta L_{(d-s)}^*$, $\Delta C_{ab(d-s)}^*$, and $\Delta H_{ab(d-s)}^*$, while the Pearson correlation test is carried out between $\Delta E_{(d-s)}^*$ and $\Delta L_{(d-s)}^*$, $\Delta C_{ab(d-s)}^*$, and $\Delta H_{ab(d-s)}^*$, as shown in Table 4. The results of Pearson correlation analysis showed that $\Delta E_{(d-s)}^*$ was negatively correlated with $\Delta L_{(d-s)}^*$ and $\Delta C_{ab(d-s)}^*$, and positively correlated with $\Delta H_{ab(d-s)}^*$, and the correlation with $\Delta L_{(d-s)}^*$ was the most significant. That is, within a certain range, the larger the $\Delta L_{(d-s)}^*$ or the $\Delta C_{ab(d-s)}^*$, the smaller the $\Delta E_{(d-s)}^*$. The larger the $\Delta H_{ab(d-s)}^*$, the larger the $\Delta E_{(d-s)}^*$.

Through Pearson correlation, it is possible to determine the close relationship between the color difference of double-sided heterochromatic digital printing and single-sided printing, lightness difference, and chroma difference. In order to investigate whether the law between the objective color difference, the lightness, chroma, and hue angle of the experimental sample can be expressed by ΔL_{d-s}^* , ΔC_{d-s}^* , and $\Delta H_{ab(d-s)}^*$, the regression analysis between $\Delta E_{(d-s)}^*$ and $\Delta L_{(d-s)}^*$, $\Delta C_{ab(d-s)}^*$, and $\Delta H_{ab(d-s)}^*$ is carried out, and it can be found that there is a linear correlation between them, and the regression equation can be expressed as follows:

Objective color difference range

$$\Delta E_{(d-s)}^* = 5.884 - 0.698 \Delta L_{(d-s)}^* - 0.114 \Delta C_{ab(d-s)}^* + 0.497 \Delta H_{ab(d-s)}^* (R^2 = 0.854) \quad (2)$$

where ΔL_{d-s}^* is the lightness difference between double-sided and single-sided printing, ΔC_{d-s}^* is the chroma difference between double-sided and single-sided printing, $\Delta H_{ab(d-s)}^*$ is the hue angle difference between double-sided and single-sided printing. Significance analysis $R^2 = 0.85$ demonstrates that the model fits well and may be used to forecast the color difference range while doing double-sided heterochromatic digital printing. The results indicate that when producing double-sided heterochromatic digital printing on chiffon fabric, there will be a large color difference between the single-sided and the double-sided printing, and the effect is greatly affected by $\Delta L_{(d-s)}^*$ and $\Delta H_{ab(f-b)}^*$, and is least affected by $\Delta C_{ab(d-s)}^*$. In conclusion, the model can predict the degree of

color difference between single-sided and double-sided digital printing according to the $L_{ab}^* C_{ab}^* h_{ab}$ value of the specified color.

In order to evaluate the performance of the regression models (1) and (2), six testing colors were selected. The $L^* a^* b^*$ values of these colors are shown in Supplementary Table 2, and the six colors were paired in pairs to generate a total of 30 heterochromatic double-sided experimental samples. According to the analysis method of the above experiment, the $\Delta E_{(f-b)}^*$ and $\Delta E_{(d-s)}^*$ of the 30 colors are calculated, and the experimental data of these 30 colors are substituted into the above two regression equations, respectively. Then, the Pearson correlations between $\Delta E_{(f-b)}^*$ and $\Delta E_{f-b(\text{regression prediction})}^*$, and between $\Delta E_{(d-s)}^*$ and $\Delta E_{d-s(\text{regression prediction})}^*$ were analyzed. The Pearson correlation coefficient between $\Delta E_{(f-b)}^*$ and $\Delta E_{f-b(\text{regression prediction})}^*$ is 0.50, and the Pearson correlation coefficient between $\Delta E_{(d-s)}^*$ and $\Delta E_{d-s(\text{regression prediction})}^*$ is 0.90. Therefore, it can be seen that the size of the color difference simulated by the regression equation is closely related to the actual size of the color difference between samples and that both can be used to simulate the actual color difference, among which the simulation performance of the model (2) is better than that of the model (1).

Analysis of subjective evaluation of double-sided heterochromatic digital printing

In the subjective experiment, the participants were asked to visually score the double-sided heterochromatic digital printing under only visual influence, with the degree of the rating representing the perceived color difference. A Pearson correlation analysis was conducted between visual rating and color lightness, chroma, and hue angle in order to provide a model of subjective color difference range for double-sided digital printing and to discuss the color characteristics within the subjectively acceptable color difference range for double-sided digital printing on chiffon fabric.

Observer repeatability

Before evaluating the validity of the test results, it is necessary to examine the repeatability and precision of the observer test. The main reasons for visual matching errors include visual fatigue caused by long-term observation, psychological factors of observers during the matching process, and fewer repetitions of visual matching experiments. In order to assure the correctness of the assessment test results, the volunteers were instructed to manage their time efficiently so that the subjective experiment could be finished within 30 min. Figure 6 depicts a box plot of the visual ratings of the initial

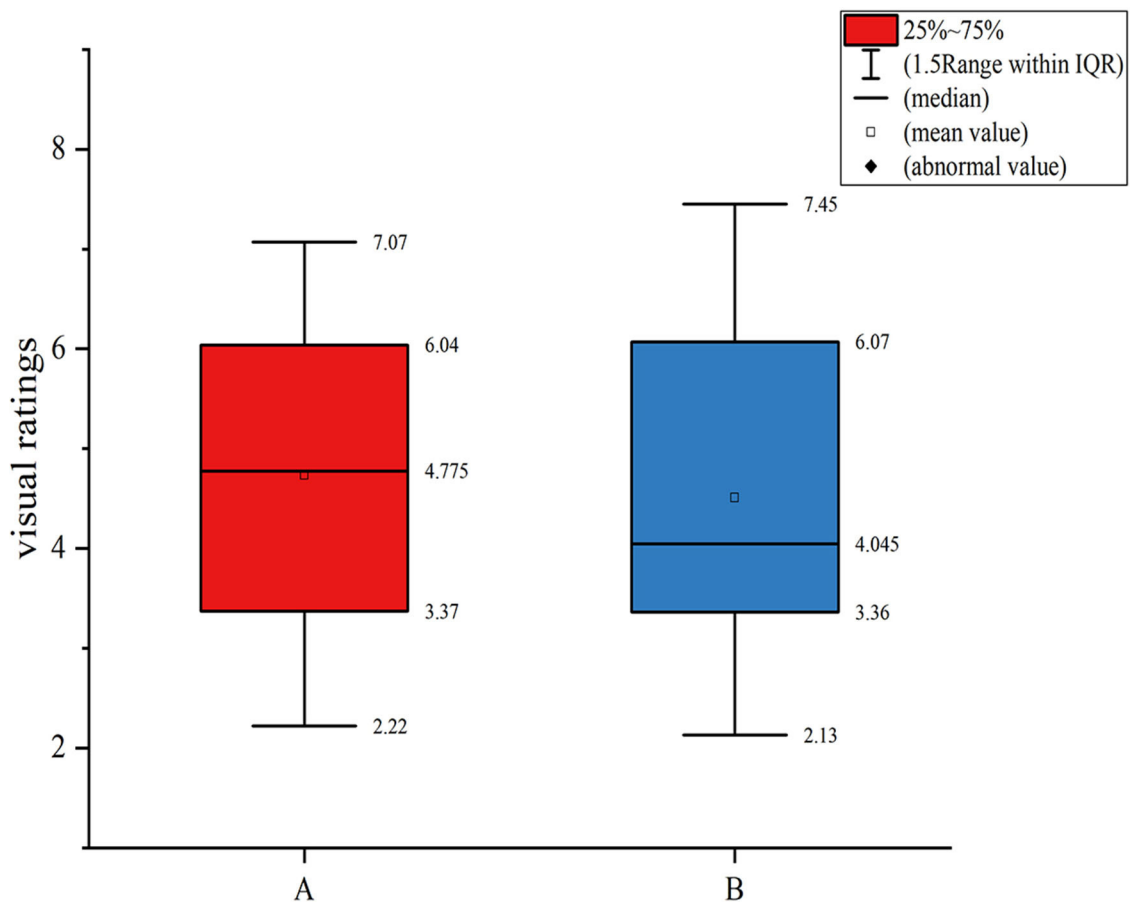


FIGURE 6
Boxplot of 30 subjects' visual scores.

experiment data A and the repeated experiment data B, revealing that there were no outliers in the subjective ratings of the 30 participants, suggesting that the subjective ratings of all 30 subjects were utilized for data analysis.

Two measures were used to examine the reliability of the experimental data: inter-observer variability and intra-observer variability. The former indicates how well the observers agreed with each other in visual responses. For each observer, the inter-observer variability was determined by the root mean square (RMS) value:

$$\text{RMS} = \sqrt{\frac{\sum_i x_i - \bar{x}}{N}} \quad (3)$$

Here x_i represents the visual score for observers "i" to the sample, \bar{x} represents the mean visual score of all observers to the sample, and N is the number of the sample. The lower the RMS value, the more closely the observers agreed with each other. Inter-observer variability, however, is concerned with how repeatable is each observer's response. This was also determined by Equation (3), with \bar{x} being replaced by y_i , the response

of observer "i" at the second occurrence of the sample. To evaluate intra-observer variability, all samples were presented two times for each observer. The lower the RMS value, the more repeatable is the observer's response. Table 5 shows the inter-observer and intra-observer variability for all observers, while the mean value for female observers was lower than for male observers, indicating that female observers were more consistent and repeatable.

Subjective color difference law of double-sided heterochromatic digital printing

Following the analysis of the relationship between objective color difference and chroma, lightness, and hue, Pearson correlation analysis was performed between the subjective scores and L^*_{ab} , C^*_{ab} , h_{ab} values of these 210 experimental samples in order to investigate the relationship between the intensity

TABLE 5 Inter-observer and intra-observer variability for experiment.

Male	Inter-observer variability	Intra-observer variability	Female	Inter-observer variability	Intra-observer variability
Observer 1	1.17	1.64	Observer 1	2.18	1.74
Observer 2	2.25	1.38	Observer 2	0.66	1.58
Observer 3	3.29	1.64	Observer 3	1.94	2.08
Observer 4	2.21	1.95	Observer 4	1.36	1.27
Observer 5	2.39	2.84	Observer 5	2.23	3.30
Observer 6	2.42	1.62	Observer 6	2.50	3.56
Observer 7	2.19	1.98	Observer 7	1.56	1.37
Observer 8	1.33	1.85	Observer 8	3.13	2.23
Observer 9	1.74	2.32	Observer 9	1.91	1.95
Observer 10	2.07	1.75	Observer 10	1.33	1.50
Observer 11	1.90	1.93	Observer 11	1.42	1.80
Observer 12	2.53	1.91	Observer 12	2.38	1.68
Observer 13	2.25	1.62	Observer 13	1.72	1.08
Observer 14	1.36	1.64	Observer 14	2.26	1.52
Observer 15	2.05	2.27	Observer 15	1.70	1.36
Male mean	2.08	1.89	Female mean	1.88	1.87

TABLE 6 Pearson correlation between perceptual color difference and L^* , C_{ab}^* , and h_{ab} .

	L^*	C_{ab}^*	h_{ab}
Pearson correlation coefficient	−0.18**	−0.48**	0.10
Sig.	0.008	0.00	0.145

**Indicates that the test method used in this paper is a two-tailed significance test method.

of subjective color difference, the lightness, and the chroma of color during double-sided heterochromatic digital printing on translucent chiffon fabric and the analysis results as shown in Table 6.

The findings demonstrate a substantial Pearson correlation between the size of the perceived color difference and the L^* and C_{ab}^* (Sig<0.01), suggesting that the model is extremely significant. The Pearson correlation with hue angle was not statistically significant. The Pearson correlation coefficients for perceived color difference with L^* and C_{ab}^* were−0.18 and−0.48, respectively, indicating a negative correlation between perceived color difference and lightness and chroma; the smaller the L^* and C_{ab}^* of the designed color within a certain range, the larger the perceived color difference, and in double-sided digital printing, the objective color difference between single- and double-sided digital printing is strongly influenced by lightness and chroma. Meanwhile, using Pearson correlation, a significant association (Sig<0.01) between subjective color difference and objective color difference (correlation coefficient of 0.84) was found. It shows that the

perceived color difference under subjective assessment and the objective color difference under instrumental measurement have a positive connection and are closely associated, with the lower the color difference, the lower the perceived color difference rating. To some extent, it can be considered that when the subjective color difference is <3, the objective color difference is acceptable. The intensity of objective color difference may be utilized to determine whether the color difference of this sample is subjectively acceptable by humans in practical production applications.

Subjective color difference range of double-sided heterochromatic digital printing

The prediction model of the color difference range based on the objective measurement of the instrument is presented in the preceding research on the objective color difference range of double-sided heterochromatic digital printing. To determine whether the subjective color difference when doing double-sided heterochromatic digital printing can be directly predicted by the relationship between the lightness, chroma, and hue angle of the color, and whether there is a difference between men and women when making the rating (Zhang et al., 2019), the differences in $L^*C_{ab}^*h_{ab}$ between the single- and double-sided experimental samples were denoted as $\Delta L_{(d-s)}^*$, $\Delta C_{ab(d-s)}^*$, and $\Delta H_{ab(d-s)}^*$, and the Pearson correlation analyses were conducted between the mean values of subjective and combined subjective scores of men and women, $\Delta L_{(d-s)}^*$, $\Delta C_{ab(d-s)}^*$, and $\Delta H_{ab(d-s)}^*$,

TABLE 7 Pearson correlation coefficient.

	$\Delta L^*_{(d-s)}$	$\Delta C^*_{ab(d-s)}$	$\Delta H^*_{ab(d-s)}$
Pearson correlation between visual scores of females and $\Delta L^*_{(d-s)}$, $\Delta C^*_{ab(d-s)}$, $\Delta H^*_{ab(d-s)}$			
Pearson correlation coefficient	-0.79**	-0.02	0.74**
Sig.	0.000	0.764	0.000
Pearson correlation between visual scores of males and $\Delta L^*_{(d-s)}$, $\Delta C^*_{ab(d-s)}$, $\Delta H^*_{ab(d-s)}$			
Pearson correlation coefficient	-0.75**	-0.095	0.72**
Sig.	0.000	0.155	0.000
Pearson correlation between visual scores of comprehensive visual scores and $\Delta L^*_{(d-s)}$, $\Delta C^*_{ab(d-s)}$, $\Delta H^*_{ab(d-s)}$			
Pearson correlation coefficient	-0.77**	-0.07	0.735**
Sig.	0.000	0.333	0.000
	Male	Female	Comprehensive
Pearson correlation between physical color difference $\Delta E^*_{(d-s)}$ and male, female, comprehensive visual score			
Pearson correlation coefficient	0.80**	0.86**	0.84**
Sig.	0.000	0.000	0.000

**Indicates that the test method used in this paper is a two-tailed significance test method.

as shown in Table 7. To investigate whether the magnitude of the visual scores was influenced by gender, Pearson correlation analysis was done between the objective color difference $\Delta E^*_{(d-s)}$ (men and women), as well as the combined visual scores under the instrumental measurements, as shown in Table 7. The Pearson correlation analysis results showed that the correlation between the subjective scores as well as the combined subjective scores of men and women and the lightness difference between single- and double-sided printing was significant, with women having a higher correlation than men. In the correlation analysis between subjective ratings and objective color difference for both genders, the correlation between subjective ratings and objective ratings measured by the instrument was also higher for women than for men, and female observers responded more accurately to color than male observers. This finding is consistent with the findings of Yang and Baroun's team (Baroun and Alansari, 2006; Yang and Li, 2016).

The Pearson correlation shows that in the subjective assessment of double-sided heterochromatic digital printing, there is a close correlation between the perceived color difference (visual rating), the difference in lightness, and the hue between single and double sides. In order to investigate whether the law of the perceived color difference can be expressed by $\Delta L^*_{(d-s)}$,

$\Delta C^*_{ab(d-s)}$, and $\Delta H^*_{ab(d-s)}$, the regression analysis of the mean values of objective scores, the composite objective scores of men and women, $\Delta L^*_{(d-s)}$, $\Delta C^*_{ab(d-s)}$, and $\Delta H^*_{ab(d-s)}$ was conducted, which revealed a linear correlation between them, and the regression equation can be expressed as follows:

Subjective prediction of female visual score

$$F = 1.375 - 0.131\Delta L^*_{(d-s)} + 0.001\Delta C^*_{ab(d-s)} + 0.117\Delta H^*_{ab(d-s)} (R^2 = 0.754) \quad (4)$$

Subjective prediction of male visual score

$$M = 2.545 - 0.137\Delta L^*_{(d-s)} - 0.012\Delta C^*_{ab(d-s)} + 0.122\Delta H^*_{ab(d-s)} (R^2 = 0.709) \quad (5)$$

Subjective prediction of composite visual score

$$C = 1.956 - 0.135\Delta L^*_{(d-s)} - 0.006\Delta C^*_{ab(d-s)} + 0.119\Delta H^*_{ab(d-s)} (R^2 = 0.743) \quad (6)$$

where $\Delta L^*_{(d-s)}$ is the lightness difference between double-sided and single-sided printing, $\Delta C^*_{(d-s)}$ is the chroma difference between double-sided and single-sided printing, and $\Delta H^*_{ab(d-s)}$ is the hue angle difference between double-sided and single-sided printing. Based on the significance analysis of the regression equation, the regression model fitting degree among female visual scores and $\Delta L^*_{(d-s)}$, $\Delta C^*_{ab(d-s)}$, and $\Delta H^*_{ab(d-s)}$ is the highest ($R^2 = 0.75$), and the regression model fitting degree between male visual score and $\Delta L^*_{(d-s)}$, $\Delta C^*_{ab(d-s)}$, as well as $\Delta H^*_{ab(d-s)}$ is the lowest ($R^2 = 0.71$), while the regression model fitting degree between combined visual score and $\Delta L^*_{(d-s)}$, $\Delta C^*_{ab(d-s)}$, and $\Delta H^*_{ab(d-s)}$ is $R^2 = 0.74$, and a scatter plot between visual results and model predictions is shown in Figure 7.

When executing double-sided heterochromatic digital printing, this model could be used to forecast the range of perceived color differences. When combined with Pearson correlation analysis, it is possible to infer that lightness has the biggest influence on subjective rating in the subjective evaluation, which is consistent with the results of Xiaoming Zhao's research teams (Zhao et al., 2014). From this model, it can be concluded that the subjective color difference is greatly affected by $\Delta L^*_{(d-s)}$ and $\Delta H^*_{ab(d-s)}$, and is less affected by $\Delta C^*_{(d-s)}$. This conclusion is consistent with the objective color difference model (2), but the subjectively acceptable color difference range is larger than that of the objective color difference. In conclusion, when carrying out the color design of double-sided digital printing of different colors, the lightness difference needs to be considered first, followed by the hue difference.

When provided subjective measures of colors, female observers reacted more correctly to colors than males, according

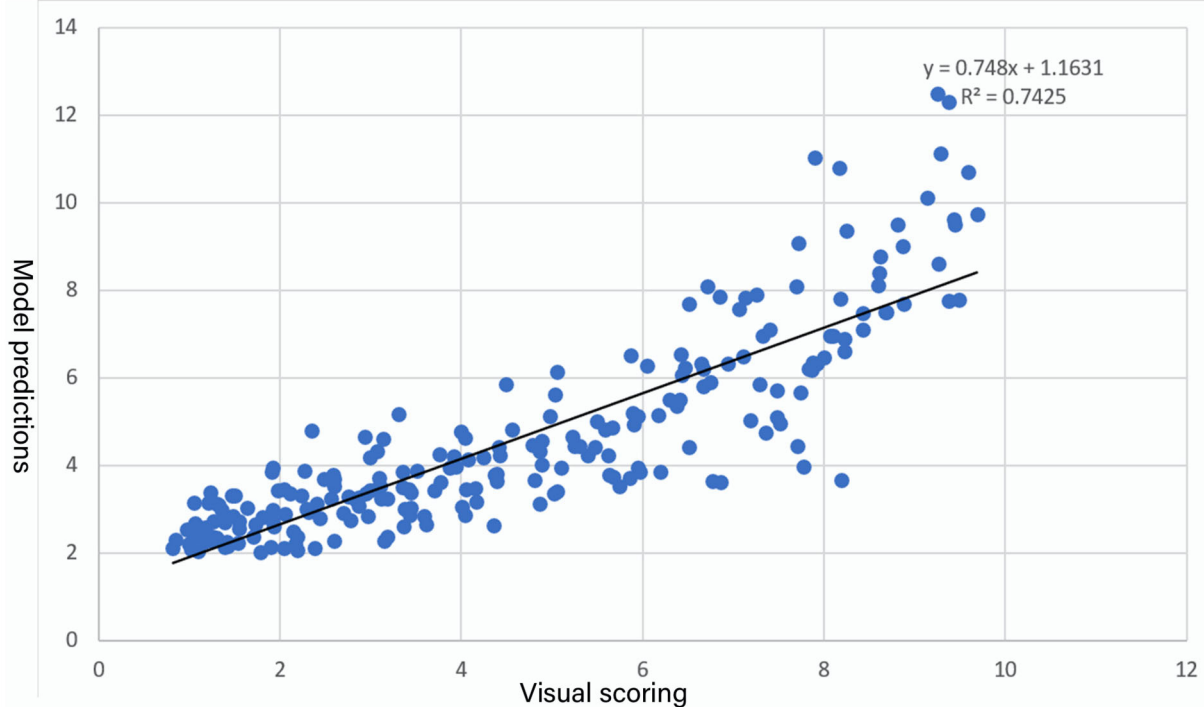


FIGURE 7
Scatter plot between visual scoring and model predictions.

to an analysis of the regression models under the visual evaluations from subjective males and females. Moreover, it was discovered that the combined mean score provided by women was 3.95, and the combined mean score supplied by men was 5.21 by assessing the magnitude of the visual ratings for both genders. This leads to the conclusion that in this trial, women were more tolerant to color than men.

Color distribution within the range of subjectively acceptable color difference

It is known from the related research that the experimental samples with objective color difference <3 in textiles are acceptable (Liu et al., 2007). The 210 groups of double-sided heterochromatic digital printing experimental samples selected for this experiment have only nine groups of objective color difference <3 , accounting for 4.2%, proving that when doing double-sided heterochromatic digital printing on chiffon fabric, it will generally produce a large color difference, so there is a limitation in using the above objective experimental results model. The appraisal of the printing quality needs to be based on human feelings,

with subjective human judgment being trustworthy. According to the scale of subjective experiment rules, the subjectively acceptable color difference in this experiment is <3 .

The subjective color difference of 210 groups of double-sided heterochromatic digital printing experimental samples chosen for this research was <3 , with 78 groups accounting for 37.1%, and the physical color difference at this time is all within 15. It demonstrates that when making double-sided heterochromatic digital printing on chiffon fabric, the range of subjective acceptable color difference to human sight exceeds the actual color difference. To visualize the distribution of chromaticity values of the experimental samples in CIELAB color space, the two-dimensional distribution of a^* - b^* when the color difference is <3 is shown in Figure 8A. These colors are not distributed in quadrant 3; the values of a^* and b^* in the first and second quadrants can exceed 50, while the distribution below 20 is less and even close to zero. Values of $|a^*|$ in the fourth quadrant are in the (10, 30) range, whereas $|b^*|$ values are in the (20, 40) interval. As demonstrated in Figure 8B for the distribution of 78 colors in CIELAB space, there is no color dispersion near the origin.

This paper modeled the L^* and a^* , b^* in MATLAB to more intuitively analyze the color characteristics within the

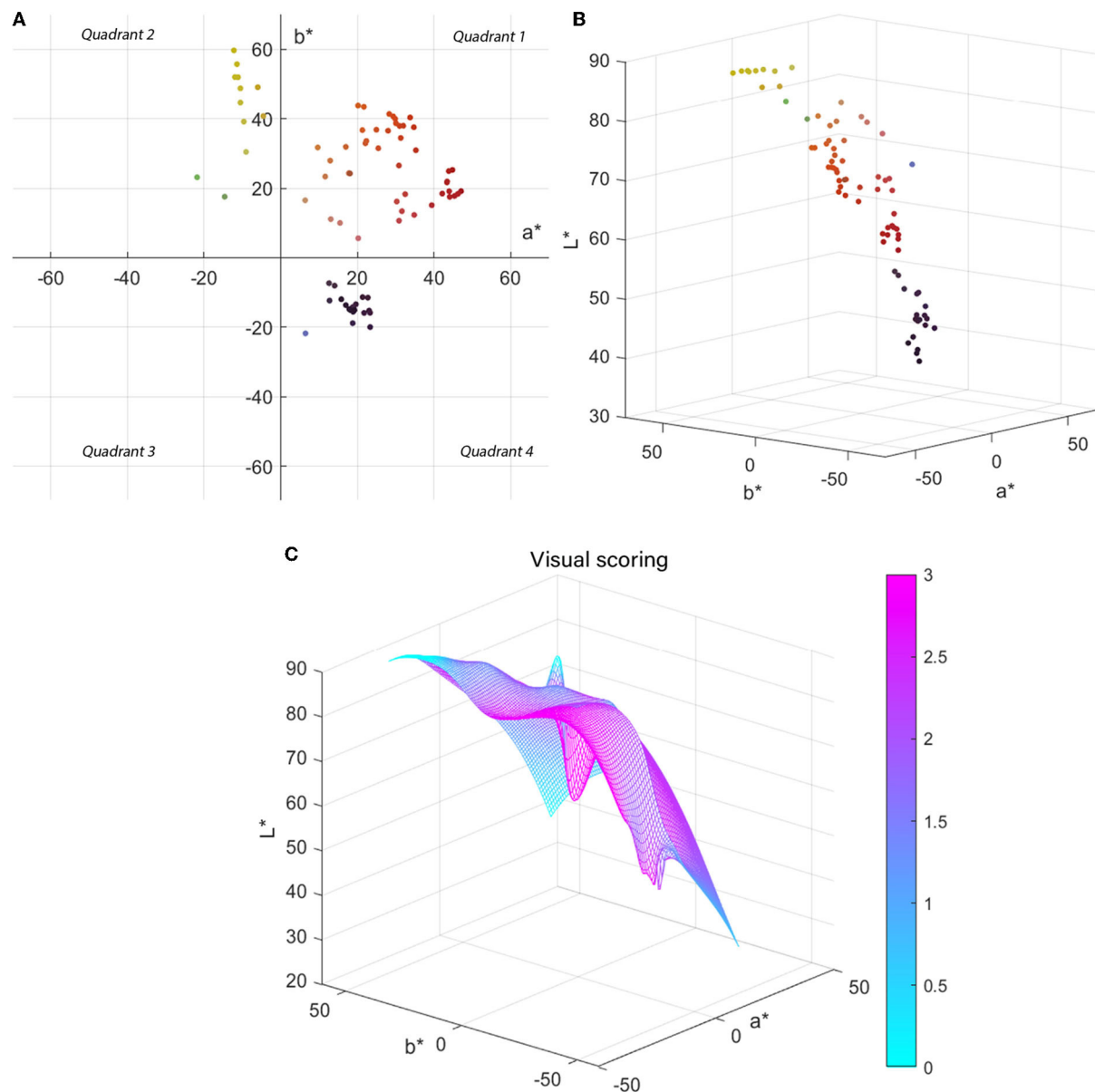


FIGURE 8
(A) Two-dimensional distribution of 78 colors in CIELAB space. **(B)** Three-dimensional distribution of 78 colors in CIELAB space. **(C)** Correlation model of chromaticity value $L^*a^*b^*$ and human visual score.

subjective acceptable color difference range during double-sided heterochromatic digital printing, as shown in Figure 8C, with colors indicating the relationship between the $L^*a^*b^*$ values of the fabric and the subjective visual rating; the colors from blue to pink indicate the deepening of the influence of $L^*a^*b^*$ values on subjective visual scores.

In conclusion, based on an analysis of the distribution of chromaticity values $L^*a^*b^*$ in CIELAB color space with a subjective visual score of <3 , when it comes to double-sided heterochromatic digital printing on chiffon fabric, the

colors within the subjective acceptable color difference have the following characteristics: when the color is in the first quadrant, the lightness ranges between 55 and 75, and the chromatic value a^*b^* ranges between 15 and 55. When the color is in the second quadrant, the lightness is between 20 and 65, the chromatic value a^* is between -15 and 0 , and b^* , which appears to be bigger, is between 20 and 60. When the color is in the fourth quadrant, the chromatic value is between 30 and 45, and the chromatic value a^*b^* is between 10 and 25.

Conclusion and prospects

Conclusion

Based on the combination of subjective evaluation of psychophysics experiment and objective instrument measurement, the main goal of this paper is to analyze the relationship between the subjective color difference and objective color difference and lightness, chroma, and hue of the color when it comes to double-sided heterochromatic digital printed on chiffon fabrics. At the same time, this research is to provide a prediction model of the dominant color determination for double-sided heterochromatic digital printing. The dominant color model has a wide range of applications. It can help the designers to understand when heterochromatic color pairs are mixed on the front and back, and the color rendered will be closer to whichever color, i.e., which color is more dominant. The prediction models of the range of color differences created under subjective and objective measurements are offered *via* the research of subjective and objective experimental data. Through the combination usage of the objective color difference model and the subjective color difference model, people can understand whether the printing effect of the designed color is acceptable to the human eye during the double-sided digital printing process, so as to improve the color matching problem of double-sided digital printing with heterochromatic colors. This model is suitable for translucent natural fibers printed with activated fuels, and the key conclusions of this research are as follows.

From the objective experiment, the following conclusions can be derived

- (1) Based on the analysis of color difference intensity of the front and back of double-sided digital heterochromatic printing, it can be seen that color performance tends to be closer to the lighter color in double-sided heterochromatic digital printing. So, in the actual production, one needs to choose a pair of colors with similar lightness for double-sided digital printing. The prediction model for the dominant color determination is as follows:

Dominant color determination

$$\Delta E^*_{(f-b)} = -0.102 + 4.554 \Delta L^*_{(f-b)} - 1.084 \Delta C^*_{ab(f-b)} + 0.024 \Delta H^*_{ab(f-b)} (R = 0.745) \quad (1)$$

(When $\Delta E^*_{f-b} < 0$, the positive color is the dominant color)

- (2) The objective color difference of double-sided heterochromatic digital printing is closely related to the lightness and chroma of the color and shows a negative correlation; the objective color difference range of double-sided heterochromatic digital printing is affected by the

lightness difference between single- and double-sided digital printing and the chroma difference, which is the most affected by the lightness difference. The prediction model of the objective color difference range is as follows.

Objective color difference range

$$\Delta E^*_{(d-s)} = 5.884 - 0.698 \Delta L^*_{(d-s)} - 0.114 \Delta C^*_{ab(d-s)} + 0.497 \Delta H^*_{ab(d-s)} (R = 0.854) \quad (2)$$

From the subjective measurement experiments, the following conclusions can be derived

- (1) The perceived color difference in double-sided heterochromatic digital printing is negatively related to the lightness and chroma of the color. The lightness difference between single- and double-sided digital printing influences the range of perceived color difference in double-sided heterochromatic digital printing. The color difference range prediction model is as follows.

Subjective prediction of composite visual score

$$C = 1.956 - 0.135 \Delta L^*_{(d-s)} - 0.006 \Delta C^*_{ab(d-s)} + 0.119 \Delta H^*_{ab(d-s)} (R = 0.743) \quad (6)$$

- (2) When men and women make subjective assessments of experimental samples, female observers are more color tolerant than males, but the fitting degree of the regression equation of the color difference range between men and women shows that female observers' experimental results are closer to the objective results than shown by male observers.
- (3) According to the correlation model between L^* , a^* , b^* and subjective visual score for experimental samples with subjective measurement < 3 , when doing double-sided heterochromatic digital printing, the color with higher lightness (55–75) in the first quadrant and chromaticity value a^*b^* between 15 and 55 need to be chosen. Colors in the second quadrant with lightness values between 20–65 and smaller b^* value (0–15) can be matched in pairs with colors in the fourth quadrant with lightness values between 30–45 and smaller a^*b^* values (10–25).

Given the objective experiment and subjective measurement experiment, the following conclusions can be drawn

The law of color presentation under subjective measurement is closely connected to the law of color presentation under objective measurement (Pearson correlation coefficient is 0.84). When selecting colors for double-sided heterochromatic digital

printing, lightness has the greatest impact on color difference, followed by the hue difference.

Prospects

By combining subjective assessment of psychophysics experiment and objective instrument measurement, the effects of color lightness, chroma, and color relative to color rendering were studied during double-sided heterochromatic digital printing. The associated prediction model is also provided. In terms of color selection, the chosen colors are unevenly distributed in CIELAB space, and the colors in the third quadrant are absent, and therefore the sample size of the selected colors may be increased in future trials. Because the subjective experiment is an observation of the combined effect of perceived lightness, chroma, and hue, the experimental results are closely related to the subjects' own color learning ability and the experimental conditions, and the experiment conducted in a dark room included only subjects from university, with no experimental samples from other age groups. Alternative backdrop circumstances and the inclusion of additional age groups in future tests need to be attempted. This experiment was limited to translucent chiffon fabric, which is thinner and has a greater permeability than other silk fabrics, and the color rendering impact was influenced by the higher permeability in the chiffon fabric sample. In future studies, alternative colors can be tested to see if the pattern design has an effect on the color rendering. With the ongoing growth of color technology and the expansion in human life's desire for color in apparel, the research of the color presentation pattern for double-sided heterochromatic digital printing of textiles has a broad range of possibilities.

Meaning of symbols

In the paper, all terms of the dominant color model, objective color difference model, and subjective color difference model are shown in [Supplementary Table 3](#).

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary material](#), further inquiries can be directed to the corresponding authors.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in

this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

MZ and YD contributed to the conception and design of the study, performed the statistical analysis, and wrote the manuscript. MZ and HY performed the experimental part. All authors contributed to the revision, read, and approved the submitted version.

Funding

This work was supported by the Youth Project of Zhejiang Provincial Natural Science Foundation (No. LQ19C090009), and the Research Fund of Zhejiang Sci Tech University (Nos: 19012201-Y, 18012108-Y, and XMJWCb20200029). The authors declare that this study received funding from Honghua Digital Technology Co., Ltd. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Conflict of interest

Authors HZ and LX are currently employed by Hangzhou Honghua Digital Technology Co., Ltd. Author YD is doing postdoctoral cooperation research in Hangzhou Honghua Digital Technology Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer PR declared a shared affiliation with the author YD to the handling editor at the time of review.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2022.956748/full#supplementary-material>

References

- Baroun, K., and Alansari, B. (2006). Gender differences in performance on the stroop test. *Soc. Behav. Pers.* 34, 309–318. doi: 10.2224/sbp.2006.34.3.309
- Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods.* 39, 175–191. doi: 10.3758/BF03193146
- Gangakhedkar, N. S. (2010). Colour measurement methods for textiles. *Colour Meas.* 10, 221–252. doi: 10.1533/9780857090195.2.221
- Gooby, B. (2020). The development of methodologies for color printing in digital inkjet textile printing and the application of color knowledge in the ways of making project. *J. Text. Des. Res. Pract.* 8, 358–383. doi: 10.1080/20511787.2020.1827802
- Liu, H. X., Huang, M. in., and Wu, B., Xu, Y. F. (2007). Subjective evaluation of chromatic aberration in color printing images. *J. Beijing Inst. Graph. Des.* 2007, 1–4. doi: 10.19461/j.cnki.1004-8626.2007.02.001
- Luo, M. R., Cui, G., and Rigg, B. (2001). The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color Res. Appl.* 26, 340–350. doi: 10.1002/col.1049
- Ou, L. C., Luo, M. R., Sun, P. L., Hu, N. C., and Chen, H. S. (2012). Age effects on colour emotion, preference, and harmony. *Color Res. Appl.* 37, 92–105. doi: 10.1002/col.20672
- Qin, Y. (2014). Chinese colors. *Text. Sci. Res.* 4, 56–57. Available online at: <https://kns.cnki.net/kcms/detail/detail.aspx?FileName=FZKX201404030&DbName=CJFQ2014>
- Tyler, D. J. (2005). Textile digital printing technologies. *Text. Prog.* 37, 1–65. doi: 10.1533/tepr.2005.0004
- Tyler, D. J. (2011). Digital printing technology for textiles and apparel. *Comput. Technol. Text. Apparel.* 12, 259–282. doi: 10.1533/9780857093608.3.259
- Wang, H., Cui, G., Luo, M. R., and Xu, H. (2012). Evaluation of colour-difference formulae for different colour-difference magnitudes. *Color Res. Appl.* 37, 316–325. doi: 10.1002/col.20693
- Wang, J. H., and Dou, M. (2005). Standardization of color management (II). *Mod. Paint Coating* 02, 50–51+54. Available online at: <https://kns.cnki.net/kcms/detail/detail.aspx?FileName=XDTL200502018&DbName=CJFQ2005>
- Wang, Y., Zhou, H., Mei, Z., Zhou, T., and Yang, J. Z. (2016). Evaluation of total ink volume control effect in textile digital printing. *Silk* 53, 36–42. Available online at: <https://kns.cnki.net/kcms/detail/detail.aspx?FileName=SICO201602007&DbName=CJFQ2016>
- Westland, S., and Pan, Q. (2017). Advances in instrumental colour pass/fail analysis. *J. Text. Sci. Eng.* 7, 6. doi: 10.4172/2165-8064.1000321
- Yang, F., and Li, C. (2016). The color of gender stereotyping: The congruity effect of topic, color, and gender on health messages' persuasiveness in cyberspace. *Comput. Hum. Behav.* 64, 299–307. doi: 10.1016/j.chb.2016.07.001
- Yang, X. H., Sui, J. H., Bin, M., and Zhang, X. W. (2010). Auto-generating uniform stochastic web images for ink-jet printing textiles. *Text. Res. J.* 80, 1942–1948. doi: 10.1177/0040517510373637
- Zhang, Y., Liu, P., Han, B., Xiang, Y., and Li, L. (2019). Hue, chroma, and lightness preference in Chinese adults: Age and gender differences. *Color Res. Appl.* 44, 967–980. doi: 10.1002/col.22426
- Zhao, X., Teng, P., and Zong, J. (2014). Research on the ability of human eyes to distinguish color difference of different colors. *Electron. Sci. Technol.* 1, 303–307. doi: 10.16453/j.issn.2095-8595.2014.03.001



OPEN ACCESS

EDITED BY

Alyssa A. Brewer,
University of California, Irvine,
United States

REVIEWED BY

Shu Zhao,
Anhui University, China
Hamidreza Bolhasani,
Islamic Azad University, Iran

*CORRESPONDENCE

Linquan Lv
linquanlv@163.com

†These authors have contributed
equally to this work

SPECIALTY SECTION

This article was submitted to
Perception Science,
a section of the journal
Frontiers in Neuroscience

RECEIVED 13 July 2022

ACCEPTED 02 November 2022

PUBLISHED 24 November 2022

CITATION

Lv L, Peng M, Wang X and Wu Y (2022)
Multi-scale information fusion
network with label smoothing
strategy for corneal ulcer
classification in slit lamp images.
Front. Neurosci. 16:993234.
doi: 10.3389/fnins.2022.993234

COPYRIGHT

© 2022 Lv, Peng, Wang and Wu. This is
an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in
other forums is permitted, provided
the original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Multi-scale information fusion network with label smoothing strategy for corneal ulcer classification in slit lamp images

Linquan Lv^{1*†}, Mengle Peng^{2†}, Xuefeng Wang¹ and
Yuanjun Wu¹

¹Anhui Finance and Trade Vocational College, Hefei, Anhui, China, ²Department of Mechanical and Energy Engineering, Zhejiang University of Science and Technology, Hangzhou, Zhejiang, China

Corneal ulcer is the most common symptom of corneal disease, which is one of the main causes of corneal blindness. The accurate classification of corneal ulcer has important clinical importance for the diagnosis and treatment of the disease. To achieve this, we propose a deep learning method based on multi-scale information fusion and label smoothing strategy. Firstly, the proposed method utilizes the densely connected network (DenseNet121) as backbone for feature extraction. Secondly, to fully integrate the shallow local information and the deep global information and improve the classification accuracy, we develop a multi-scale information fusion network (MIF-Net), which uses multi-scale information for joint learning. Finally, to reduce the influence of the inter-class similarity and intra-class diversity on the feature representation, the learning strategy of label smoothing is introduced. Compared with other state-of-the-art classification networks, the proposed MIF-Net with label smoothing achieves high classification performance, which reaches 87.07 and 83.84% for weighted-average recall (W_R) on the general ulcer pattern and specific ulcer pattern, respectively. The proposed method holds promise for corneal ulcer classification in fluorescein staining slit lamp images, which can assist ophthalmologists in the objective and accurate diagnosis of corneal ulcer.

KEYWORDS

corneal ulcer classification, multi-scale information fusion, label smoothing, deep learning, fluorescein staining slit lamp images

Introduction

Corneal ulcer is a serious blinding eye disease, which is one of the main causes of corneal blindness (Smith, 2004; Deswal et al., 2017; Lopes et al., 2019). In addition, it is an inflammatory or more serious infectious corneal disease involving disruption of the stroma epithelial layer, which can cause great pain to the patient and may cause severe

vision loss or even blindness (Cohen et al., 1987; Diamond et al., 1999; Manikandan et al., 2019).

Standardized screening, timely diagnosis and early treatment are effective ways to reduce the blindness rate of corneal ulcer. Fluorescent staining is often used to observe the integrity of the ocular surface, especially the integrity of the cornea. Therefore, fluorescent staining technology has become a common tool to assist ophthalmologists in diagnosing corneal ulcer, and can provide great convenience for the diagnosis and treatment of corneal ulcer (Kaufman, 1960; Schweitzer, 1967; Morgan and Carole, 2009; Kumar and Thirumalesh, 2013; Zhang et al., 2018). The current corneal ulcer prevention and treatment model is based on the ulcer's general pattern and type grade (TG) standards, using slit lamp microscopy combined with fluorescent staining technology to examine the ocular surface of high-risk groups (Wolffsohn and Purslow, 2003; Khanal et al., 2008; Peterson and Wolffsohn, 2009). For the ulcer's general pattern, the diagnostic procedure for corneal ulcer is based on the shape and distribution characteristics of the corneal ulcer, which can be classified into three categories, respectively, corresponding to point-like corneal ulcers, point-flaky mixed corneal ulcers and flaky corneal ulcers (Smith, 2004; Deswal et al., 2017). While the TG grading method has two components: (1) according to the specific pattern of corneal ulcers, it is divided into five categories (type0–type4), which is usually the first step for ophthalmologists to diagnose underlying disease. (2) according to the location of corneal ulcer in the cornea, it is divided into five categories (grade0–grade4). The use of fluorescent stained images to identify corneal ulcers plays an important role in formulating treatment plans. However, due to differences in subjective experience and professional knowledge, ophthalmologists have certain differences in the recognition of corneal ulcers based on fluorescent stained images. Therefore, it is very important to study an objective and accurate automatic classification method of corneal ulcers, which can assist ophthalmologists in formulating individualized drug or surgical intervention strategies.

In the past, many related works on automated or semi-automated methods for corneal ulcers are mainly for the segmentation of corneal ulcers, which is considered to be a pixel-by-pixel classification. For example, Wolffsohn and Peterson et al. applied the color extraction algorithm and edge detection algorithm of the RGB system to automatically segment the ulcer area (Wolffsohn and Purslow, 2003; Peterson and Wolffsohn, 2009), while Pritchard et al. (2003) used threshold technology to indirectly detect conjunctival hyperemia and punctate corneal ulcers. Chun et al. (2014) used RGB and hue-saturation-value (HSV) techniques to evaluate corneal staining on 100 images. Deng et al. (2018b) successively used k-means clustering, morphological operations, and region growth to achieve automatic ulcer segmentation, and they also proposed a simple linear iterative clustering (SLIC) based super-pixel

method (Deng et al., 2018a). In addition, Liu et al. used the combined method of Otsu and Gaussian mixture modeling (GMM) to segment the intracorneal ulcer area on 150 images. In recent years, deep learning has received widespread global attention, and its automated analysis of ophthalmic images has also made a huge breakthrough (Liu et al., 2018; Peng et al., 2020, 2021, 2022a,b,c; Wang et al., 2020). Recently, some progress has been made in the detection and segmentation of corneal ulcers based on deep learning. For example, Sun et al. (2017) proposed a patch-based deep convolutional neural network to segment corneal ulcers. A recent study proposed a system for automatically detecting corneal ulcer disease (Akram and Debnath, 2019), which first uses the Haar cascade classifier to detect and segment the eye part of the face, and then, the convolutional neural network (CNN) is used to detect the presence of corneal ulcer disease. If there is corneal ulcer, active contour technology is used to locate and segment the ulcer area.

The above studies mainly focus on the segmentation of corneal ulcer. Few studies involve the classification of corneal ulcer, which is an important reference for ophthalmologists to formulate treatment strategies. Therefore, the current paper builds upon the previous successful models and proposes a simple and effective methodology, which combines multi-scale information fusion and label smoothing strategy to achieve two classification patterns of corneal ulcer. The first is to realize the classification of corneal ulcer according to the ulcer's general pattern, while the second is to achieve the type grading of corneal ulcers according to the specific ulcer pattern. The relevant images of corneal ulcer are shown in the **Figure 1**, and a detailed description of the relevant symptoms and the number of fluorescein staining images in each category is given in **Table 1**. It can be seen from **Figure 1** and **Table 1** that it is challenging to achieve accurate corneal ulcer classification mainly due to the following two reasons: (1) Corneal ulcer has complex pathological features and noise interference; (2) Different types of corneal ulcers are similar in pathological shape and distribution. In this study, for point-like corneal ulcers, point-flaky mixed corneal ulcers and type1 corneal ulcers with micro punctate, continuous down sampling of the pool layer in CNN may lead to the loss of lower resolution features related to category, which may lead to the decline of classification performance. Previous studies have shown that some shallow features can improve classification performance (Lee et al., 2015; Wang et al., 2015; Ma et al., 2019), which has attracted our attention. In addition, considering that the misclassification of corneal ulcers may be caused by images of similar but different categories in the dataset (Wan et al., 2021), we introduce label smoothing in the cross-entropy loss, which can reduce the inter-class similarity and intra-class differences (Müller et al., 2019). In short, we develop a multi-scale information fusion network with label smoothing strategy to achieve the classification of corneal ulcer. The main contributions of this paper can be summarized as follows:

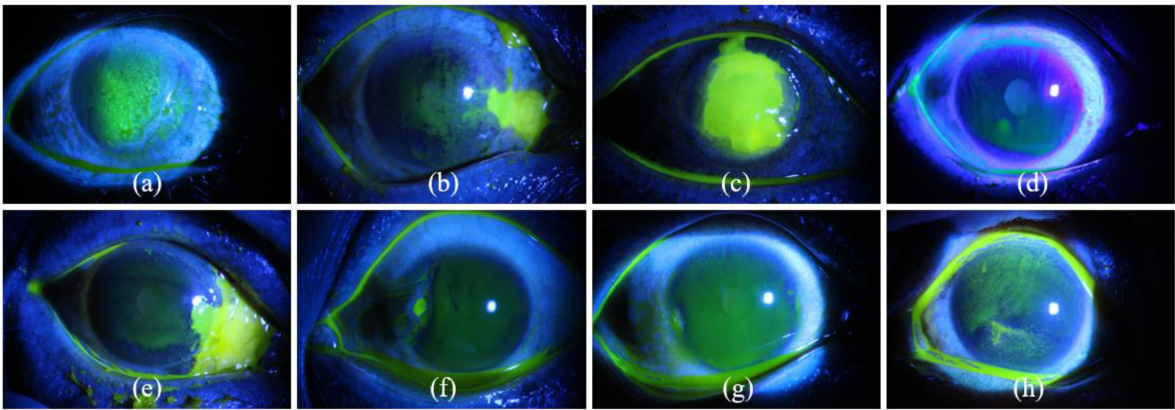


FIGURE 1
Eight examples of slit lamp images of corneal ulcers. Panels (a–c) represent point-like corneal ulcer, point-flaky mixed corneal ulcer, and flaky corneal ulcer according to general ulcer pattern, while panels (d–h) are type0, type1, type2, type3, and type4 corneal ulcers according to specific ulcer pattern.

TABLE 1 Symptoms and number of different types of corneal ulcers.

Ulcer pattern	Categories	Symptoms	Num	Total
General ulcer pattern	Point-like corneal ulcers	A large number of small ulcers are gathered and can be distributed anywhere on the cornea.	358	712
	Point-flaky mixed corneal ulcers	Between punctate and lamellar corneal ulcers, there are both punctate and lamellar ulcers in the cornea with irregular distribution.	263	
	Flaky corneal ulcers	The ulcer area is usually bright green with clear borders.	91	
Specific ulcer pattern	Type0	No ulcer of the corneal epithelium.	36	712
	Type1	Micro punctate.	78	
	Type2	Macro punctate.	40	
	Type3	Coalescent macro punctate.	10	
	Type4	Patch (>=1 mm).	548	

“Num” and “Total” represent the number of fluorescein staining images in each category and total number of fluorescein stained images.

- (1) To make full use of shallow edge information and deep semantic information, a multi-scale information fuser is designed to enhance feature expression capabilities, which can improve the robustness of prediction results.
- (2) Label smoothing strategy is introduced to reduce the impact of the inter-class similarity and intra-class differences between corneal ulcer images on feature representation and guide the model to learn salient features with category differences.
- (3) Extensive experiments are conducted to evaluate the effectiveness of the proposed MIF-Net and the results show that the proposed MIF-Net outperforms other state-of-the-art classification networks.

The remainder of this paper is organized as follows: The proposed method for automatic corneal ulcer classification is introduced in Section 2. Section 3 presents the experimental results in detail. In section 4, we conclude this paper and suggest future work.

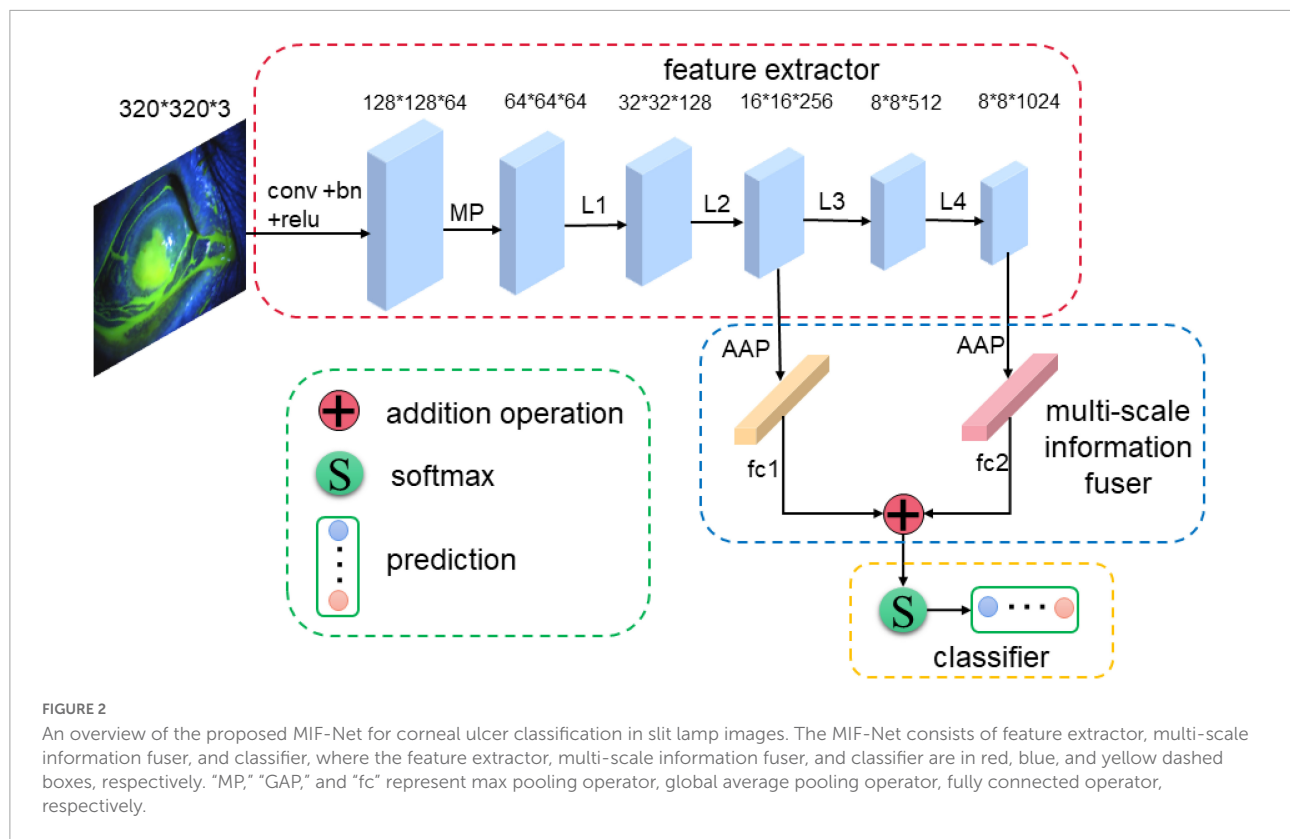
Materials and methods

Overview

The proposed MIF-Net for corneal ulcer classification is shown in **Figure 2**, which consists of three main parts: feature extractor, multi-scale information fuser, and classifier with label smoothing. Firstly, the feature extractor is used to extract spatial features from the input cornea ulcer image. Then, the multi-scale information fuser is adopted to fuse the classification information from the two different layers of feature extractor. Finally, the label smoothing strategy is applied to the final corneal ulcer classification.

Feature extractor

To achieve corneal ulcer classification, a backbone network needs to be constructed for feature extraction. We have



firstly compared several common backbone networks in image classification, including ResNet (He et al., 2016), DenseNet (Huang et al., 2017), Inception (Ioffe and Szegedy, 2015; Szegedy et al., 2016, 2017), EfficientNet (Tan and Le, 2019), ResNext (Xie et al., 2017), VGGNet (Simonyan and Andrew, 2014). In addition, the above backbone networks have been studied and analyzed in detail. Firstly, experimental results in Tables 2, 3 show the performance of ResNet50 (He et al., 2016), DenseNet121 (Huang et al., 2017), InceptionV3 (Szegedy et al., 2016), ResNext50 (Xie et al., 2017), and VGG16 (Simonyan and Andrew, 2014) is better and comparable. Secondly, compared with ResNet50 (He et al., 2016), InceptionV3 (Szegedy et al., 2016), ResNext50 (Xie et al., 2017), and VGG16 (Simonyan and Andrew, 2014), DenseNet121 (Huang et al., 2017) uses dense connections to encourage features reuse, reduce the number of network parameters greatly, where each front layer function is used as the input of the latter layer. Therefore, considering the balance of model performance and simplicity, we choose DenseNet121 as the backbone network for feature extraction, where the global average pooling layer and the fully connected layer in original version are removed. In addition, many previous studies have shown transfer learning is an effective strategy to speed up the training convergence and improve the classification performance (Pan and Yang, 2010; He et al., 2016; Huang et al., 2017; Tan et al., 2018). Therefore, transfer learning is used to help model training in this study.

Multi-scale information fuser

As we know, deep networks have brought performance gains in computer vision tasks, which can extract global feature with deep semantic information. However, many empirical evidences suggest that the performance improvement cannot be achieved by simply stacking more layers (Shen et al., 2016). In addition, most of convolutional neural networks (CNNs) based image classification methods only use the final global feature extracted by the feature extractor for classification, and the shallow local feature does not directly participate in the network training, which may lead to the neglect of some important details for the classification of corneal ulcer. In addition, the current classification, detection and segmentation networks usually use convolutional neural networks to extract the characteristics of objects through layer by layer abstraction. The deeper the network is, the larger the receptive field is, the stronger the semantic information representation ability is, and it is suitable for processing large objects (Chen et al., 2017; George et al., 2018). However, the lower the resolution of the feature map, the weaker the geometric information representation ability is. In the shallow layer of the network, the smaller the down sampling multiple, the smaller the receptive field is, the ability to represent geometric details is strong, and it is suitable for processing small targets (Zhao et al., 2017; George et al., 2018). Although the resolution is high, the ability to represent semantic information

TABLE 2 The results of comparable experiments on general classification of corneal ulcers.

Methods	W_R (%)	W_P (%)	W_F1 (%)	Kappa (%)	Parameters (M)
ResNet18 (He et al., 2016)	83.28 ± 2.49	83.20 ± 2.25	83.00 ± 2.43	76.72 ± 3.88	11.1781
ResNet34 (He et al., 2016)	83.98 ± 2.75	84.06 ± 2.76	83.80 ± 2.78	76.84 ± 1.20	21.2862
ResNet50 (He et al., 2016)	84.26 ± 2.31	84.32 ± 2.01	84.00 ± 2.21	78.03 ± 3.49	23.5142
DenseNet169 (Huang et al., 2017)	80.33 ± 2.31	70.58 ± 1.84	75.03 ± 1.90	66.15 ± 7.36	12.4895
InceptionResNetV2 (Ioffe and Szegedy, 2015)	83.98 ± 2.51	83.69 ± 2.54	83.56 ± 2.57	74.66 ± 2.17	54.3111
InceptionV3 (Szegedy et al., 2016)	84.68 ± 1.91	84.42 ± 2.09	84.19 ± 1.90	74.84 ± 4.82	21.7917
InceptionV4 (Szegedy et al., 2017)	82.29 ± 2.05	82.32 ± 2.07	82.19 ± 1.99	74.20 ± 1.67	41.1472
ResNxt50 (Xie et al., 2017)	84.82 ± 2.19	84.67 ± 2.16	84.57 ± 2.04	78.63 ± 1.94	22.9861
SE_ResNet50 (Hu et al., 2018)	85.96 ± 2.40	86.83 ± 0.92	86.56 ± 0.97	77.50 ± 1.76	26.0452
SE_ResNext50 (Hu et al., 2018)	83.71 ± 1.08	83.33 ± 1.44	83.03 ± 1.11	74.35 ± 2.11	25.5170
VGG16 (Szegedy et al., 2016)	84.69 ± 1.96	84.34 ± 2.07	84.38 ± 1.95	79.06 ± 3.57	134.2637
EfficientNetB2 (Ioffe and Szegedy, 2015)	81.88 ± 2.24	81.64 ± 2.77	80.87 ± 2.83	72.91 ± 4.91	7.7095
EfficientNetB4 (Ioffe and Szegedy, 2015)	81.75 ± 1.94	81.70 ± 2.47	80.99 ± 2.57	73.19 ± 6.14	17.5594
LmNet (Wan et al., 2021)	85.52 ± 2.24	85.64 ± 2.34	85.17 ± 2.25	78.71 ± 3.25	23.6875
Backbone	84.39 ± 4.05	84.38 ± 4.28	84.25 ± 4.24	78.42 ± 6.31	6.9569
Proposed	87.07 ± 1.98	86.93 ± 2.15	86.82 ± 2.07	81.49 ± 3.25	6.9577

Bold values indicate the best performance.

TABLE 3 The results of comparable experiments on specific classification of corneal ulcers.

Methods	W_R (%)	W_P (%)	W_F1 (%)	Kappa (%)	Parameters (M)
ResNet18 (He et al., 2016)	81.88 ± 1.82	74.82 ± 3.94	77.76 ± 2.76	53.23 ± 9.08	11.1781
ResNet34 (He et al., 2016)	80.19 ± 1.38	72.85 ± 2.95	75.17 ± 3.14	44.15 ± 23.62	21.2862
ResNet50 (He et al., 2016)	81.18 ± 0.51	72.98 ± 1.73	76.49 ± 0.79	54.25 ± 5.25	23.5142
DenseNet169 (Huang et al., 2017)	80.77 ± 1.41	73.24 ± 2.09	75.84 ± 2.39	64.95 ± 10.54	12.4895
InceptionResNetV2 (Ioffe and Szegedy, 2015)	81.88 ± 1.57	76.32 ± 5.42	78.26 ± 3.79	53.06 ± 16.65	54.3111
InceptionV3 (Szegedy et al., 2016)	81.74 ± 1.93	77.52 ± 3.75	79.13 ± 2.59	63.11 ± 8.05	21.7917
InceptionV4 (Szegedy et al., 2017)	81.74 ± 1.11	78.22 ± 3.95	79.30 ± 2.40	63.00 ± 10.40	41.1472
ResNxt50 (Xie et al., 2017)	82.02 ± 2.18	77.66 ± 4.31	78.95 ± 3.72	62.29 ± 9.21	22.9861
SE_ResNet50 (Hu et al., 2018)	81.46 ± 2.58	76.98 ± 4.16	78.71 ± 3.36	57.47 ± 12.53	26.0452
SE_ResNext50 (Hu et al., 2018)	81.87 ± 1.94	78.25 ± 6.68	79.04 ± 4.25	54.99 ± 21.31	25.5170
VGG16 (Szegedy et al., 2016)	81.74 ± 0.56	75.69 ± 4.08	78.17 ± 2.10	57.20 ± 14.30	134.2637
EfficientNetB2 (Ioffe and Szegedy, 2015)	81.17 ± 1.87	73.05 ± 2.45	76.30 ± 2.28	51.04 ± 10.19	7.7095
EfficientNetB4 (Ioffe and Szegedy, 2015)	81.88 ± 1.48	75.18 ± 4.80	78.12 ± 3.34	58.37 ± 11.86	17.5594
LmNet (Wan et al., 2021)	82.42 ± 0.89	74.68 ± 2.30	78.23 ± 1.09	61.60 ± 8.97	23.6875
Backbone	81.45 ± 1.48	74.88 ± 4.10	77.28 ± 2.32	55.39 ± 12.17	6.9569
Proposed	83.84 ± 1.61	78.32 ± 2.26	80.52 ± 2.25	72.06 ± 6.75	6.9577

Bold values indicate the best performance.

is weak. Based the above theory and analysis, the challenge for corneal ulcer classification in this study that different categories of corneal ulcers present lesions of different shapes and sizes can just correspond to the deep and shallow features of the convolution neural network. For example, for point-like corneal ulcers and point-flaky mixed corneal ulcers and type1 corneal ulcers with micro punctate, the feature map is continuously down-sampled through the pooling layer in the CNN, which may lead to the loss of lower resolution features related to the category, so it is very important to retain shallow feature

information. For flaky corneal ulcers and type2, type3, type4 corneal ulcers with macro punctate or larger patch, the high-level semantic information is more conducive to improving the classification accuracy due to its relatively large target size. Therefore, inspired by the deep supervised learning and multi-scale feature fusion strategy, which can combine the shallow edge information and deep semantic information to improve the classification performance, we design a new multi-scale information fuser to improve performance with a small increase in the number of model parameters, which is shown in [Figure 2](#).

As can be seen from **Figure 2**, the feature maps of the second and fourth stages of DenseNet121 are fed to two global average pooling layers, respectively, and then the fully connected layers are used to convert them into probability distributions. Finally, we fuse these two predictions to generate the final prediction as shown in Equation 1. Therefore, predictions generated from different levels can supervise the model training.

$$P = f(P_1, P_2) \quad (1)$$

where P_1 and P_2 are the prediction results of the second and fourth stages of DenseNet121, P is the corresponding information fusion results. In addition, $f(\cdot)$ denotes the information fusion operation.

Label smoothing strategy

As can be observed from **Figure 1** and **Table 1**, corneal ulcer images have the characteristics of large intra-class diversity and high inter-class similarity to a certain extent. The traditional cross-entropy loss function only calculates the loss that the predicted value is the real class, which may lead to poor classification performance of the model in this study due to inter-class similarity and intra-class difference. Therefore, we introduce label smoothing strategy into the cross-entropy loss function to reduce the similarity between classes and difference within classes and improve the generalization of the model, which can alleviate the overconfidence problem caused by the traditional cross-entropy loss function.

Suppose D is a classification dataset with M samples (x_i, y_i) ($i = 1, 2, \dots, M$), where x_i and y_i represent an input image and its corresponding category label, respectively. A standard multi-classification problem is to predict the probability of the input image x_i , belonging to category k ($y_i = k$). The category k is encoded by one-hot labels as a vector $t = (0, 0, \dots, 0, 1, 0, \dots, 0)$, where only t_k is 1 and all others are 0. This characteristic encourages the model to learn in the direction with the greatest difference between the correct label and the wrong label, which means only the loss of the correct label position is calculated in the optimization process of the model. However, when the training data is small, and the inter-class similarity and intra-class differences is relatively large, it may cause the network to be overfitting (Wan et al., 2021). To solve the above problems and inspired by previous study (Szegedy et al., 2017; Müller et al., 2019), label smoothing is introduced in this study. It is a regularization strategy, which mainly adds noise through soft one-hot to reduce the weight of the true label category and slightly increases the penalty for the wrong label category in the training of the model, and ultimately reduces the risk of overfitting. The label smoothing strategy used in this study is as follows:

$$t' = t * (1 - \varepsilon) + I * \frac{\varepsilon}{N} \quad (2)$$

where t and t' represent the one-hot labels before and after label smoothing. ε is a random number between 0.1 and 0.2, which can be regarded as noise introduced in the fixed distribution. I is a matrix with the same dimension as t , and its element values are all one. N is the total number of categories.

Loss function

In this study, we propose a multi-scale information fusion network (MIF-Net), which takes the original cornea ulcer images as input. In addition, considering the similarity between classes and differences within classes, label smoothing strategy is used in network optimization. Based on the analysis, the cross-entropy loss function based on label smoothing strategy is used and defined as follow:

$$L = -\frac{1}{m} \sum_{i=1}^m \sum_{k=1}^K f(t_i = k) * \log(p(k | x_i)) + (1 - f(t_i = k)) * \log(1 - p(k | x_i)) \quad (3)$$

where m is the number of samples in per mini-batch, t_i is the class label of the input image x_i . $f(\cdot)$ is an indicator function, which is one if t_i equals k ($k = 1, 2, \dots, K$).

Experiments and results

In this section, we first introduce the experimental dataset in detail. Then, the experimental setup will be described, including imaging processing, the parameter settings in the training phase and evaluation metrics in the testing phase. Finally, we will give the detailed experimental results and the corresponding analysis.

Dataset

In this study, the SUSTech-SYSU public dataset (Deng et al., 2020) is used to evaluate the proposed MIF-Net, which contains a total of 712 fluorescein staining images with ground truth annotated in image-wise by three experienced ophthalmologists from Zhongshan Ophthalmic Center at Sun Yat-sen University. The fluorescein staining image with a resolution of 2592×1728 pixels contains only one cornea, which is completely presented in the image, roughly in the center of the visual field. The labeling of corneal ulcers classification is based on the symptoms described in **Table 1**. It can be seen from **Table 1** that the category distribution is unbalanced, where most cornea ulcer data are point-like corneal ulcers, point-flaky mixed corneal ulcers, and type4 corneal ulcers and the data of several other corneal ulcers is relatively few. To evaluate the effectiveness of the proposed MIF-Net, a 5-fold cross-validation strategy is

adopted, where the whole dataset is randomly divided into 5 subsets of the almost same size according to the proportion of the number of each category.

Experimental setup

Image processing

To reduce the computational cost and improve the computational efficiency of the model, all the images are resized to 320×320 by bilinear interpolation and normalized to [Lopes et al. \(2019\)](#). In addition, online data augmentation, including random rotation 30° , horizontal flipping, and vertical flipping, is adopted to prevent over-fitting and improve the robust ability of the model.

Parameter setting

The proposed MIF-Net is performed on the public platform Pytorch. We use A NVIDIA GTX Titan X GPU with 12GB memory to train the model with back-propagation algorithm by minimizing the loss function as illustrated in Equation 3. The Adam was used as the optimizer, where both initial learning rate and weight decay are set to 0.0001. The batch size and epoch are set to 16 and 50, respectively. To ensure fairness, all the networks in this study are trained with same optimization schemes and we save the best model on validation set. The code of the proposed MIF-Net will be released in: <https://github.com/linquanlv0915/MIF-Net>.

Evaluation metrics

Considering the category imbalance of the dataset shown in [Table 1](#) and to comprehensively and fairly evaluate the classification performance of different methods, four common evaluation indicators are used ([Peng et al., 2021, 2022b](#)), including weighted-average recall (W_R), weighted-average precision (W_P), weighted-average F1 score (W_F1), Kappa index ([McHugh, 2012](#)).

Results

In this study, we propose a classification network named MIF-Net with label smoothing for two classification patterns of corneal ulcer, including general ulcer pattern and specific ulcer pattern. In our experiments, we evaluate performance of the proposed method on the 712 fluorescein staining images by using 5-fold cross-validation strategy, which randomly divided the whole dataset into 5 subsets according to the proportion of the number of each category. In each experiment, model was trained with 4 subsets and test on the remaining one subset. The experiments were repeated 5 times with each of the 5 subsets used exactly once as the testing set and the other four subsets as training set and the final experimental results were averaged over all the experiments. Next, a series

of comparison experiments and ablation experiments are presented and analyzed in detailed. For convenience, the basic DenseNet121 pretrained on ImageNet is called Backbone.

Results on general ulcer pattern

As can be seen from [Table 1](#), the corneal ulcers can be divided into three categories according to the general ulcer pattern, namely point-like corneal ulcers, point-flaky mixed corneal ulcers and flaky corneal ulcers. [Table 2](#) shows the quantitative results of different methods. We compare the proposed method with other excellent CNN based classification networks, including ResNet18 ([He et al., 2016](#)), ResNet34 ([He et al., 2016](#)), ResNet50 ([He et al., 2016](#)), DenseNet169 ([Huang et al., 2017](#)), InceptionResNetV2 ([Ioffe and Szegedy, 2015](#)), InceptionV3 ([Szegedy et al., 2016](#)), InceptionV4 ([Szegedy et al., 2017](#)), ResNext50 ([Xie et al., 2017](#)), SE_ResNet50 ([Hu et al., 2018](#)), SE_ResNext50 ([Hu et al., 2018](#)), VGG16 ([Simonyan and Andrew, 2014](#)), EfficientNetB2 ([Tan and Le, 2019](#)), and EfficientNetB4 ([Tan and Le, 2019](#)). It can be seen from [Table 2](#) that our method achieves superior performance in term of all evaluation metrics.

First, compared with Backbone, the performance of the proposed method has been greatly improved, which improves the W_R, W_P, W_F1, and Kappa by 3.92, 3.02, 3.05, and 3.91%, respectively, and achieves 87.07% for W_R, 86.93% for W_P, 86.82% for W_F1, and 81.49% for Kappa. Then, compared with other state-of-the-art classification networks, our MIF-Net with label smoothing gets an overall improvement in terms of all indicators with comparable or less model complexity. For example, compared with the best performance among the comparison classification networks (SE_ResNet50), the proposed method with less model parameters improves the W_R, W_P, W_F1, and Kappa by 1.29, 0.12, 0.30, and 5.15%, respectively. In addition, compared with EfficientNetB2 with comparable model parameters, our proposed method has also made great improvement in terms of all evaluation metrics. It is worth noting that the proposed method is also compared with a recent study on multi-scale feature fusion and label smoothing, which is proposed for remote sensing classification and is named LmNet ([Wan et al., 2021](#)). LmNet takes pre-trained ResNext50 as backbone and combines channel attention, multi-scale feature fusion and label smoothing. As can be seen from [Table 2](#), the proposed MIF-Net with label smoothing outperforms LmNet on all metrics, which improves the W_R by 1.75%. Moreover, the model complexity of our method is less than that of LmNet. These results demonstrate the effectiveness of the proposed method in the general classification pattern of corneal ulcers.

Results on specific ulcer pattern

It can be seen from [Table 1](#), the corneal ulcers can be divided into five categories according to the specific ulcer pattern, namely type0, type1, type2, type3, and type4, respectively. Similar to the general ulcer pattern, a series

TABLE 4 The results of ablation experiments on general classification of corneal ulcers.

Methods	W_R (%)	W_P (%)	W_F1 (%)	Kappa (%)	Parameters (M)
Backbone	84.39 ± 4.05	84.38 ± 4.28	84.25 ± 4.24	78.42 ± 6.31	6.9569
Backbone + Fusion_add_234	85.52 ± 2.80	85.37 ± 3.01	85.24 ± 2.87	78.69 ± 4.91	6.9592
Backbone + Fusion_add_34	85.94 ± 2.27	85.7 ± 2.02	85.50 ± 2.01	80.26 ± 1.77	6.9585
Backbone + Fusion_add_24	86.36 ± 2.70	86.71 ± 2.92	86.28 ± 2.67	80.80 ± 3.11	6.9577
Backbone + Fusion_max_234	85.25 ± 2.22	85.18 ± 2.43	84.82 ± 2.18	79.28 ± 4.34	6.9592
Backbone + Fusion_max_34	85.24 ± 2.96	85.04 ± 2.86	85.07 ± 2.93	79.92 ± 3.31	6.9585
Backbone + Fusion_max_24	84.69 ± 2.99	84.57 ± 2.80	84.49 ± 2.95	78.74 ± 6.67	6.9577
Backbone + Fusion_Concatenation_234	85.23 ± 3.61	85.15 ± 3.32	85.02 ± 3.37	80.29 ± 3.16	6.9577
Backbone + Fusion_Concatenation_34	84.96 ± 3.58	84.94 ± 3.21	84.64 ± 3.51	79.67 ± 5.97	6.9585
Backbone + Fusion_Concatenation_24	85.38 ± 3.79	84.85 ± 4.07	84.63 ± 4.17	78.58 ± 8.71	6.9592
Backbone + LS	86.21 ± 3.55	86.15 ± 3.40	86.07 ± 3.50	80.15 ± 6.49	6.9577

Bold values indicate the best performance.

TABLE 5 The results of ablation experiments on specific classification of corneal ulcers.

Methods	W_R (%)	W_P (%)	W_F1 (%)	Kappa (%)	Parameters (M)
Backbone	81.45 ± 1.48	74.88 ± 4.10	77.28 ± 2.32	55.39 ± 12.17	6.9569
Backbone + Fusion_add_24	82.44 ± 1.23	75.96 ± 2.10	79.11 ± 1.22	60.54 ± 8.74	6.9577
Backbone + LS	83.00 ± 1.29	77.68 ± 2.34	79.88 ± 1.63	65.85 ± 6.24	6.9577

Bold values indicate the best performance.

of comparison experiments with the other state-of-the-art classification networks are conducted and the quantitative results are illustrated in Table 3. As can be observed from Table 3, the performance of ResNet34 is the worst, while the performance of ResNext50 is the second best. In addition, compared to the Backbone, our proposed method gets an overall improvement in term of all evaluation indicators and improves the W_R, W_P, W_F1, and Kappa by 2.93, 4.59, 4.19, and 30.10%, respectively. It can be seen from Table 3 that the performance of the proposed method is better than other CNNs based classification networks, which achieves 83.84% for W_R, 78.32% for W_P, 80.52% for W_F1, and 72.06% for Kappa, respectively. Similarly, we also compare the proposed method with LmNet (Wan et al., 2021). As can be observed from Table 3, compared to LmNet, our proposed method achieves better performance, which improves the W_R, W_P, W_F1, and Kappa by 2.93, 3.64, 2.93, and 16.98%, respectively. The experimental results prove the effectiveness of the proposed method for the specific classification pattern of corneal ulcers in slit lamp images.

Ablation experiments

Ablation study for multi-scale information fuser

As can be seen from Figure 2, we propose a simple and effective multi-scale information strategy in this study. In this section, we explore the influence of different information fusion strategies on the corneal ulcer classification and

conduct the ablation experiments as shown in Table 4, where “Backbone + Fusion_add_234” denotes we fuse the predictions of second, third and fourth stages of DenseNet121 by addition operation to generate the final prediction and the meaning of the others is similar. In particular, “Backbone + Fusion_add_24” is our proposed fusion strategy in this study. Taking the general classification of corneal ulcers for example, it can be seen from Table 4 that compared to the Backbone, the model’s performance with all multi-scale information fusion strategies can get an overall improvement in terms of all metrics, which proves multi-scale information fusion can improve the model’s classification performance in this study. It is worth noting that the information fusion based on addition operation performs better than information fusion based on maximum operation and concatenation operation as show in Table 4. Firstly, for the corneal ulcer classification of general pattern, compared to Backbone, the performance of “Backbone + Fusion_add_24” improves by 2.33, 2.76, 2.41, and 3.03% for W_R, W_P, W_F1, and Kappa, respectively. In addition, for the corneal ulcer classification of specific pattern, the introduction of multi-scale information fusion based on addition also achieves better classification performance. As shown in Table 5, compared with the Backbone, the W_R, W_P, W_F1, and Kappa increase from 81.45, 74.88, 77.28, and 55.39% to 82.44, 75.96, 79.11, and 60.54%, respectively, which benefits from the fact that multi-scale infusion fusion can fully integrate shallow detailed information and deep semantic information to improve the multi-scale feature representation ability of the model. These

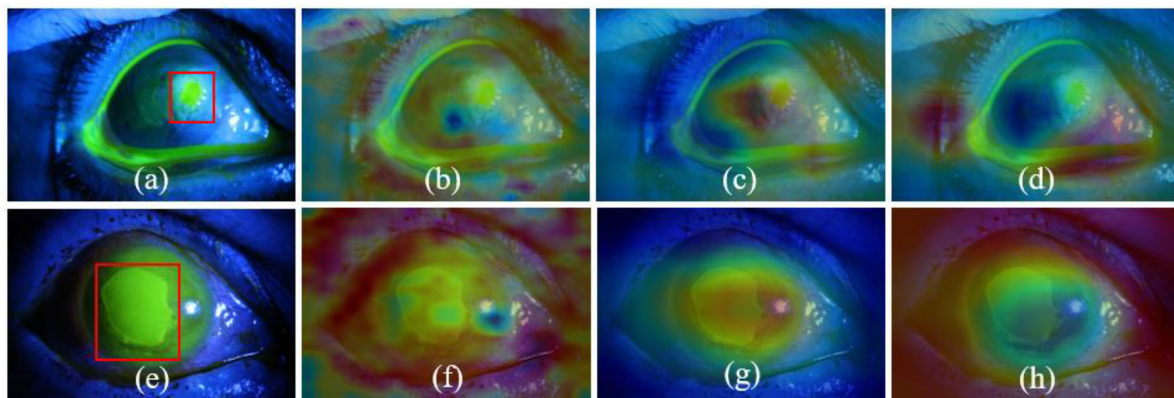


FIGURE 3

Visualization results of CAM. Panels (a,e) are the original images, where the ulcer-related pathologies are in red boxes. Panels (b,c,f,g) are the heat maps of the second and fourth stages of the proposed MIF-Net, while the panels (d,h) is the heat maps of the fourth stage of Backbone.

results indicate the effectiveness of the proposed multi-scale information fuser.

Ablation study for label smoothing strategy

It can be observed from [Tables 4, 5](#), the introduction of label smoothing strategy (Backbone + LS) also gets an overall improvement in terms of all evaluation indexes for two corneal ulcer classification tasks. Taking the corneal ulcer classification of general pattern for example, compared with the Backbone, the W_R, W_P, W_{F1}, and Kappa of the Backbone+LS increase by 2.16, 2.10, 2.16, and 2.21%, respectively, which may benefit the fact that the introduction of label smoothing can reduce the impact of similarity between classes and intra-class difference on classification performance. The results prove the effectiveness of label smoothing strategy in this study.

Conclusion and discussion

Accurate corneal ulcer classification is still a challenging task due to its complex pathology, noise interference and the similar pathological morphology and distribution of different types of corneal ulcer. In this study, to tackle these problems, a novel classification network named MIF-Net with label smoothing is proposed for corneal ulcer classification. Firstly, to avoid the loss of lower resolution features related to category caused by down sampling, a multi-scale information fuser is designed to fully integrate shallow local information and depth global features to improve the multi-scale information representation ability of the model. Then, to reduce the influence of inter-class similarity and intra-class diversity on feature representation, another learning strategy named label smoothing is developed to improve the generalization of the model, which is a modification of the loss function. The ablation experiments show that both multi-scale information fuser and label smoothing strategy

can improve the classification performance. Compared with other state-of-the-art CNN-based classification networks, the classification performance of the proposed MIF-Net with label smoothing has been improved, as shown in [Tables 2, 3](#).

To further illustrate the effectiveness of the proposed MIF-Net and increase the interpretability of CNN, we apply the “class activation mapping” technology ([Zhou et al., 2016](#)) to obtain the heat maps of fluorescein staining images with different corneal ulcer categories for the qualitative analysis, which calculates the convolutional outputs of the second and fourth stages and visualizes the class-discriminative regions concerned by the network. As can be seen from [Figures 3a,b,e,f](#), the features extracted by the proposed method in the shallow layer are local detailed information, while the features extracted from the last layer are global semantic information as shown in [Figures 3c,g](#). In addition, it can be observed from [Figures 3a,c-e,g,h](#) that compared to Backbone, our method can focus on the location of key information related to different corneal ulcer categories, which may benefit from multi-scale information fusion and label smoothing strategy. These results demonstrate that compared with Backbone, the proposed method can adaptively focus on target-related area of corneal ulcer images and efficiently improve the classification performance of corneal ulcer.

In conclusion, the proposed MIF-Net with label smoothing holds promise for corneal ulcer classification in slit lamp images. We believe that our proposed method can also be applied to other medical image classification tasks, which requires further exploration and verification. However, there is still a limitation in this study that all comparisons between the proposed method and other classification networks are based on the limited data from SUSTech-SYSU dataset. We believe that if more data are available, the performance of our method will be further improved. Therefore, in the near future, we will collect more fluorescein staining images of corneal ulcers to further evaluate the performance of the proposed method and develop relevant

semi-supervised algorithms based on MIF-Net to reduce the dependence on labeled data.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://github.com/CRazorback/The-SUSTech-SYSU-dataset-for-automatically-segmenting-and-classifying-corneal-ulcers>.

Ethics statement

The studies involving human participants were reviewed and approved by the Zhongshan Ophthalmic Centre Ethics Committee of Sun Yat-sen University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

LL designed the study, conducted most of experiments, analyzed the experimental results, and drafted the manuscript. MP conducted some of experiments, reviewed and revised

the manuscript. XW and YW reviewed the manuscript and conducted some of experiments. All authors contributed to the article and approved the submitted version.

Funding

This study was supported in part by the National Social Science Fund of China (19ZDA364) and part by the Natural Science Research Project (tzpyxj137).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Akram, A., and Debnath, R. (2019). "An efficient automated corneal ulcer detection method using convolutional neural network," in *Proceedings of the 2019 22nd international conference on computer and information technology* (Dhaka: IEEE), 1–6. doi: 10.1109/ICCIT48885.2019.9038329
- Chen, L., Papandreou, G., Schroff, F., and Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *arXiv [Preprint]*. arXiv:1706.05587.
- Chun, Y. S., Yoon, W. B., Kim, K. G., and Park, I. K. (2014). Objective assessment of corneal staining using digital image analysis. *Invest. Ophthalmol. Vis. Sci.* 55, 7896–7903. doi: 10.1167/iops.14-15618
- Cohen, E. J., Laibson, P. R., Arentsen, J. J., and Clemons, C. S. (1987). Corneal ulcers associated with cosmetic extended wear soft contact lenses. *Ophthalmology* 94, 109–114. doi: 10.1016/S0161-6420(87)33491-8
- Deng, L., Huang, H., Yuan, J., and Tang, X. (2018b). Automatic segmentation of corneal ulcer area based on ocular staining images. *Medical imaging 2018: Biomedical applications in molecular, structural, and functional imaging*. *Int. Soc. Opt. Photonics* 10578:105781D. doi: 10.1117/12.2293270
- Deng, L., Huang, H., Yuan, J., and Tang, X. (2018a). "Superpixel based automatic segmentation of corneal ulcers from ocular staining images," in *Proceedings of the IEEE 23rd international conference on digital signal processing* (Shanghai: IEEE), 1–5. doi: 10.1109/ICDSP.2018.8631795
- Deng, L., Lyu, J., Huang, H., Deng, Y., Yuan, J., and Tang, X. (2020). The SUSTech-SYSU dataset for automatically segmenting and classifying corneal ulcers. *Sci. Data* 7:23. doi: 10.1038/s41597-020-0360-7
- Deswal, J., Arya, S. K., Raj, A., and Bhatti, A. (2017). A case of bilateral corneal perforation in a patient with severe dry eye. *J. Clin. Diagn. Res.* 11, ND01–ND02. doi: 10.7860/JCDR/2017/24149.9645
- Diamond, J., Leeming, J., Coombs, G., Pearman, J., Sharma, A., Illingworth, C., et al. (1999). Corneal biopsy with tissue micro homogenisation for isolation of organisms in bacterial keratitis. *Eye* 13, 545–549. doi: 10.1038/eye.1999.135
- George, L. C., Schroff, P. F., and Adam, H. (2018). "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the conference on computer vision and pattern recognition* (Salt Lake City, UT: IEEE), 1–11.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (Las Vegas, NV: IEEE), 770–778. doi: 10.1109/CVPR.2016.90
- Hu, J., Shen, L., and Sun, G. (2018). "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (Salt Lake City, UT: IEEE), 7132–7141. doi: 10.1109/CVPR.2018.00745
- Huang, G., Liu, Z., Laurens, V. D. W., and Weinberger, K. Q. (2017). "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (Honolulu, HI: IEEE), 4700–4708. doi: 10.1109/CVPR.2017.243
- Ioffe, S., and Szegedy, C. (2015). "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the international conference on machine learning*. Mountain View, CA.
- Kaufman, H. (1960). The diagnosis of corneal herpes simplex infection by fluorescent antibody staining. *Arch. Ophthalmol.* 64, 382–384. doi: 10.1001/archophth.1960.01840010384009
- Khanal, S., Tomlinson, A., McFadyen, A., Diaper, C., and Ramaesh, K. (2008). Dry eye diagnosis. *Invest. Ophthalmol. Vis. Sci.* 49, 1407–1414. doi: 10.1167/iops.07-0635

- Kumar, A., and Thirumalesh, M. (2013). Use of dyes in ophthalmology. *J. Clin. Ophthalmol. Res.* 1, 55–58. doi: 10.4103/2320-3897.106288
- Lee, C., Xie, S., Gallagher, P., Zhang, Z., and Tu, Z. (2015). Deeply-supervised nets. *Artif. Intell. Stat.* 39, 562–570.
- Liu, H., Wong, D. W. K., Fu, H., Xu, Y., and Liu, J. (2018). “DeepAMD: Detect early age-related macular degeneration by applying deep learning in a multiple instance learning framework,” in *Proceedings of the Asian conference on computer vision* (Cham: Springer), 625–640. doi: 10.1007/978-3-030-20873-8_40
- Lopes, B. T., Eliasy, A., and Ambrosio, R. (2019). Artificial intelligence in corneal diagnosis: Where are we? *Curr. Ophthalmol. Rep.* 7, 204–211. doi: 10.1007/s40135-019-00218-9
- Ma, C., Mu, X., and Sha, D. (2019). Multi-layers feature fusion of convolutional neural network for scene classification of remote sensing. *IEEE Access* 7, 121685–121694. doi: 10.1109/ACCESS.2019.2936215
- Manikandan, P., Abdel-hadi, A., Randhir Babu Singh, Y., Revathi, R., Anita, R., Banawas, S., et al. (2019). Fungal keratitis: Epidemiology, rapid detection, and antifungal susceptibilities of *Fusarium* and *Aspergillus* isolates from corneal scrapings. *Biomed Res. Int.* 2019:6395840. doi: 10.1155/2019/6395840
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochem. Med.* 22, 276–282. doi: 10.11613/BM.2012.031
- Morgan, P. B., and Carole, M. C. (2009). Corneal staining: Do we really understand what we are seeing. *Cont. Lens Anterior Eye* 32, 48–54. doi: 10.1016/j.clae.2008.09.004
- Müller, R., Kornblith, S., and Hinton, G. E. (2019). When does label smoothing help? *arXiv [Preprint]*. arXiv:1906.02629.
- Pan, S. J., and Yang, Q. (2010). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359. doi: 10.1109/TKDE.2009.191
- Peng, Y., Chen, Z., Zhu, W., Shi, F., Wang, M., Zhou, Y., et al. (2022a). ADS-net: Attention-awareness and deep supervision based network for automatic detection of retinopathy of prematurity. *Biomed. Opt. Express* 13, 4087–4101. doi: 10.1364/BOE.461411
- Peng, Y., Chen, Z., Zhu, W., Shi, F., Wang, M., Zhou, Y., et al. (2022b). Automatic zoning for retinopathy of prematurity with semi-supervised feature calibration adversarial learning. *Biomed. Opt. Express* 13, 1968–1984. doi: 10.1364/BOE.447224
- Peng, Y., Zhu, W., Chen, Z., Shi, F., Wang, M., Zhou, Y., et al. (2022c). AFENet: Attention fusion enhancement network for optical disc segmentation of prematurity infants. *Front. Neurosci.* 13:836327. doi: 10.3389/fnins.2022.836327
- Peng, Y., Zhu, W., Chen, F., Xiang, D., and Chen, X. (2020). Automated retinopathy of prematurity screening using deep neural network with attention mechanism. *Proc. Med. Imaging* 11313, 1131321–1131327. doi: 10.1117/12.2548290
- Peng, Y., Zhu, W., Chen, Z., Wang, M., Geng, L., Yu, K., et al. (2021). Automatic staging for retinopathy of prematurity with deep feature fusion and ordinal classification strategy. *IEEE Trans. Med. Imaging* 40, 1750–1762. doi: 10.1109/TMI.2021.3065753
- Peterson, R., and Wolffsohn, J. (2009). Objective grading of the anterior eye. *Optom. Vis. Sci.* 86, 273–278. doi: 10.1097/OPX.0b013e3181981976
- Pritchard, N., Young, G., Coleman, S., and Hunt, C. (2003). Subjective and objective measures of corneal staining related to multipurpose care systems. *Cont. Lens Anterior Eye* 26, 3–9. doi: 10.1016/S1367-0484(02)00083-8
- Schweitzer, N. (1967). A fluorescein colored polygonal pattern in the human cornea. *Arch. Ophthalmol.* 77, 548–553. doi: 10.1001/archophth.1967.00980020550021
- Shen, L., Lin, Z., and Huang, Q. (2016). “Relay backpropagation for effective learning of deep convolutional neural networks,” in *Proceedings of the European conference on computer vision* (Cham: Springer), 467–482. doi: 10.1007/978-3-319-46478-7_29
- Simonyan, K., and Andrew, Z. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv [Preprint]*. arXiv:1409.1556.
- Smith, T. (2004). *BMA AZ family medical encyclopedia*. London: Dorling Kindersley Ltd.
- Sun, Q., Deng, L., Liu, J., Huang, H., Yuan, J., and Tang, X. (2017). “Patch-based deep convolutional neural network for corneal ulcer area segmentation,” in *Fetal, infant and ophthalmic medical image analysis*, eds M. Jorge Cardoso, T. Arbel, A. Melbourne, H. Bogunovic, P. Moeskops, and X. Chen (Cham: Springer), 101–108. doi: 10.1007/978-3-319-67561-9_11
- Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. (2017). “Inception-v4, Inception-resnet and the impact of residual connections on learning,” in *Proceedings of the thirty-first AAAI conference on artificial intelligence*, California, CA. doi: 10.1609/aaai.v31i1.11231
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). “Rethinking the inception architecture for computer vision,” in *Proceedings of the IEEE conference on computer vision and pattern recognition* (Las Vegas, NV: IEEE), 2818–2826. doi: 10.1109/CVPR.2016.308
- Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., and Liu, C. (2018). “A survey on deep transfer learning,” in *Proceedings of the international conference on artificial neural networks* (Berlin: Springer), 270–279. doi: 10.1007/978-3-030-01424-7_27
- Tan, M., and Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv [Preprint]*. arXiv:1905.11946.
- Wan, H., Chen, J., Huang, Z., Feng, Y., Zhou, Z., Liu, X., et al. (2021). Lightweight channel attention and multiscale feature fusion discrimination for remote sensing scene classification. *IEEE Access* 9, 94586–94600. doi: 10.1109/ACCESS.2021.3093308
- Wang, L., Chen, Y., Tu, Z., and Svetlana, L. (2015). Training deeper convolutional networks with deep supervision. *arXiv [Preprint]*. arXiv:1505.02496.
- Wang, S., Wang, X., Hu, Y., Shen, Y., Yang, Z., Gan, M., et al. (2020). Diabetic retinopathy diagnosis using multichannel generative adversarial network with semi-supervision. *IEEE Trans. Autom. Sci. Eng.* 13, 1–12.
- Wolffsohn, J., and Purslow, C. (2003). Clinical monitoring of ocular physiology using digital image analysis. *Cont. Lens Anterior Eye* 26, 27–35. doi: 10.1016/S1367-0484(02)00062-0
- Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K. (2017). “Aggregated residual transformations for deep neural networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition* (Honolulu, HI: IEEE), 1492–1500. doi: 10.1109/CVPR.2017.634
- Zhang, Y., Chen, P., Di, G., Qi, X., and Gao, H. (2018). Netrin-1 promotes diabetic corneal wound healing through molecular mechanisms mediated via the adenosine 2b receptor. *Sci. Rep.* 8:5994. doi: 10.1038/s41598-018-24506-9
- Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017). “Pyramid scene parsing network,” in *Proceedings of the IEEE conference on computer vision and pattern recognition* (Honolulu, HI: IEEE), 2881–2890. doi: 10.1109/CVPR.2017.660
- Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A. (2016). “Learning deep features for discriminative localization,” in *Proceedings of the IEEE conference on computer vision and pattern recognition* (Las Vegas, NV: IEEE), 2921–2929. doi: 10.1109/CVPR.2016.319



OPEN ACCESS

EDITED BY

Anıl Ufuk Batmaz,
Concordia University,
Canada

REVIEWED BY

Ru-Yuan Zhang,
Shanghai Jiao Tong University, China
Julia Föcker,
University of Lincoln,
United Kingdom

*CORRESPONDENCE

Kimia C. Yaghoubi
kimia.yaghoubi@email.ucr.edu

SPECIALTY SECTION

This article was submitted to
Perception Science,
a section of the journal
Frontiers in Psychology

RECEIVED 02 September 2022

ACCEPTED 31 October 2022

PUBLISHED 06 December 2022

CITATION

Yaghoubi KC, Kabbara S, Arian S, Kobaissi H,
Peters MAK and Seitz AR (2022) Comparing
random dot motion in MATLAB vs. Inquisit
Millisecond.
Front. Psychol. 13:1035518.
doi: 10.3389/fpsyg.2022.1035518

COPYRIGHT

© 2022 Yaghoubi, Kabbara, Arian, Kobaissi,
Peters and Seitz. This is an open-access
article distributed under the terms of the
Creative Commons Attribution License (CC
BY). The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Comparing random dot motion in MATLAB vs. Inquisit Millisecond

Kimia C. Yaghoubi^{1*}, Sarah Kabbara¹, Sara Arian¹, Hadi Kobaissi¹, Megan A. K. Peters^{2,3} and Aaron R. Seitz¹

¹Perception and Learning Laboratory, Department of Psychology, University of California, Riverside, Riverside, CA, United States, ²Cognitive and Neural Computation Laboratory, Department of Cognitive Sciences, University of California, Irvine, Irvine, CA, United States, ³Center for the Neurobiology of Learning and Memory, University of California, Irvine, Irvine, CA, United States

Random Dot Motion (RDM) displays refer to clouds of independently moving dots that can be parametrically manipulated to provide a perception of the overall cloud moving coherently in a specified direction of motion. As a well-studied probe of motion perception, RDMs have been widely employed to understand underlying neural mechanisms of motion perception, perceptual decision-making, and perceptual learning, among other processes. Despite their wide use, RDM stimuli implementation is highly dependent on the parameters and the generation algorithm of the stimuli; both can greatly influence behavioral performance on RDM tasks. With the advent of the COVID pandemic and an increased need for more accessible platforms, we aimed to validate a novel RDM paradigm on Inquisit Millisecond, a platform for the online administration of cognitive and neuropsychological tests and assessments. We directly compared, in the same participants using the same display, a novel RDM paradigm on both Inquisit Millisecond and MATLAB with Psychtoolbox. We found that psychometric functions of Coherence largely match between Inquisit Millisecond and MATLAB, as do the effects of Duration. These data demonstrate that the Millisecond RDM provides data largely consistent with those previously found in laboratory-based systems, and the present findings can serve as a reference point for expected thresholds for when these procedures are used remotely on different platforms.

KEYWORDS

motion perception, random dot motion, direction estimation, intra-platform validity, inter-platform validity, COVID-19

Introduction

Here, we seek to validate a novel Random Dot Motion (RDM) procedure that can be broadly accessible for research and teaching purposes. RDM refers to clouds of independently moving dots that can be parametrically manipulated to induce a perception of the overall cloud moving coherently in a specified direction of motion. RDMs provide a useful probe of motion perception, given the ability to parametrically control the relative

saliency of the motion stimuli. They have been employed in psychophysical and neurophysiological experiments to understand underlying neural mechanisms of motion perception (Britten et al., 1992; Purushothaman and Bradley, 2005), perceptual decision-making (Mazurek et al., 2003; Gold and Shadlen, 2007; Beck et al., 2008), perceptual learning (Ball and Sekuler, 1982; Zohary et al., 1994; Zanker, 1999; Seitz and Watanabe, 2003; Seitz et al., 2005; Law and Gold, 2008), motion direction discrimination (Adelson and Bergen, 1985; Cleary and Braddick, 1990; Britten et al., 1996; Banton et al., 2001; Pavan et al., 2016), depth perception (Kim and Mollon, 2002; Nadler et al., 2008; Kim et al., 2016), and short-term visual memory (Pavan et al., 2013, 2021) among other processes. While RDMs have emerged as one of the most conventional psychophysical paradigms for studying properties of visual motion perception processes, to date, RDM-based tasks are largely regulated to specialize in laboratory-based software systems (e.g., Psychophysics Toolbox for MATLAB, PsychoPy, Eprime, etc.) that are often associated with a steep learning curve, which prevents broad adaptation to students, younger researchers, clinicians, and others in early learning phases.

A challenge regarding the development of an accessible RDM-based research program is that the stimuli are somewhat complicated to generate. RDM stimuli consist of several frames in which a set of dots (often 50–200 or more) moves within an aperture, where the displacement of each dot must be calculated independently frame by frame. Further, to control for spatial cues that might indicate the direction of the dot-cloud then, when a dot exits the aperture then, a new dot needs to be generated on the opposite side of the aperture. To display such a sequence, not only does one need to generate the trajectories of the dot clouds, but one must also display them at an appropriate frame rate so that the user perceives the stimulus as intended. A further challenge is that most tasks using RDMs require these dot clouds to be generated in real-time and often in response to the performance of the participant. For instance, to estimate the RDM Coherence threshold (e.g., how well someone can perceive the motion of a small number of coherently moving dots in a cloud of otherwise randomly moving dots), it is typical for the direction, speed, and level of Coherence (e.g., the percent of dots moving in the targeted direction frame-to-frame) to change across trials, requiring a unique RDM stimulus to be generated in correspondence with adaptive procedures employed. In other words, it is not a trivial task to accurately generate or display such RDM stimuli.

However, as accentuated by the COVID pandemic, there is an increasing need for accessible platforms that can make and deploy RDM tasks, as well as other procedures, more broadly accessible and easy to use. Whether for teaching, research, or clinical use, it is important that these alternative systems can generate stimuli that have known psychometric properties. This is a particular challenge for RDMs, where the perception of RDM stimuli can be highly dependent upon both the parameters (speed, density, aperture, duration, frame rate, visual angles of dot-displacement, etc.) and the generation algorithm of the stimuli (e.g., are

randomly moving dots displaced to random new locations or do they move with a fixed speed but random direction, do dots have limited lifetimes, do the same dots move coherently frame by frame, etc.) with several studies systematically showing that precise implementation can influence behavioral performance (Watamaniuk and Sekuler, 1992; Benton and Curran, 2009; Pilly and Seitz, 2009).

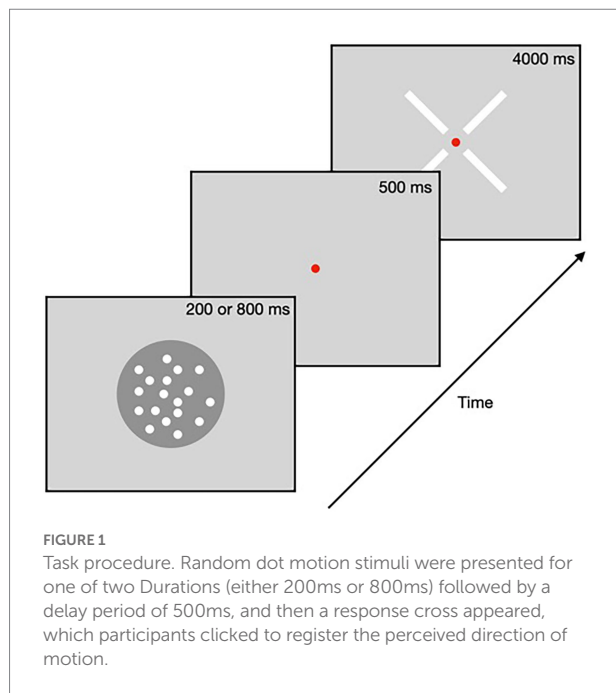
To this end, we collaborated with Millisecond, a leading provider of software platforms for the online administration of cognitive and neuropsychological tests and assessments. Inquisit Millisecond has been used in many peer-reviewed publications on a diverse range of topics, including but not limited to cognitive neuroscience, neuropsychology, clinical trials, marketing, human factors, and behavioral economics (Hillman et al., 2004; Van Damme et al., 2004, 2007; Smith et al., 2005; Vogt et al., 2008; Zakrzewska and Brzezicka, 2014; Bianchi and Laurent, 2015; Stewart et al., 2017; Jin and Lin, 2022; Smolker et al., 2022). While hundreds of experiments are programmed and tested in the Millisecond Library, until now, the RDM paradigm has not been available. In this study, we co-designed a novel RDM paradigm with Millisecond that they implemented in their Inquisit software package to closely match conventional RDMs used in MATLAB, which is implemented via PsychToolbox (Watamaniuk and Sekuler, 1992; Benton and Curran, 2009; Pilly and Seitz, 2009). The advantage of this task is that it is broadly accessible, cross-platform, and can be run on both participants' own computers and in specialized laboratory environments. Further, anyone can try out tasks for free in the Millisecond library, which makes the procedure accessible for teaching (although there are fees associated with laboratory or clinical use of the platform).

In the current study, we directly compare, in the same participants using the same display, a MATLAB-based RDM [the dotsX.m function written by Seitz and used in (Pilly and Seitz, 2009)], implemented *via* the Psychophysics Toolbox extensions version 3 (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007) compared to the novel Millisecond implementation. We find that psychometric functions of coherence largely match between the platforms, as are the effects of Duration. Further, we show that both programs have high inter-test reliability as well as high intra-test reliability between the platforms. These data demonstrate that the Millisecond RDM provides data largely consistent with those previously found in laboratory-based systems, and the present data can serve as a reference point for expected thresholds for when these procedures are used remotely on different platforms.

Materials and methods

Participants

Forty-four participants (ages 18–30) were recruited through the University of California, Riverside Psychology Research



Participation System (SONA Systems) and were compensated for their research participation by course research credit. We had targeted 50 participants as this is the standard sample size that we have used for validation studies in our lab, and recruitment was cut-off at the end of the academic quarter. All participants had normal or corrected-to-normal vision and were naive in performing the task. All participants provided informed written consent as approved by the University of California, Riverside Institutional Review Board and in accordance with the Declaration of Helsinki.

Apparatus

Participants sat on an adjustable height chair at a distance of 60 cm from a 36 cm horizontally wide ViewSonic PF817 CRT monitor set to a resolution of 1920 × 1440 and a refresh rate of 85 Hz. The distance between the participant's eyes and the monitor was fixed by having them position their head on a chin-rest. The experiment was set up such that participants' eyes and the monitor center were at the same horizontal level. Stimuli were presented *via* MATLAB, using the Psychophysics Toolbox extensions version 3 (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007) and Millisecond Inquisit Lab version 6.5.2 (Millisecond Software, LLC) using a 2015 edition Apple MacBook Pro running OS Big Sur version 11.2.3.

Stimuli

Motion stimuli consisted of RDM displays: white dots moving at a speed of 9.2 °/s on a gray background. Each dot was a

2 × 2-pixel square. A total of 100 dots were presented on every frame. Dots were displayed within an 11° diameter circular aperture that had no marked edges and was centered on the screen. Dot density was fixed across frames and trials at 16.7 dots per visual deg⁻² S⁻¹ (Pilly and Seitz, 2009). Every dot was generated to have a three-frame lifetime; at each frame transition, a set of dots moved in a coherent direction with the same speed while the rest of the dots were randomly repositioned within the aperture.

Experimental procedure

Parameters closely followed the experimental paradigm implemented by Pilly and Seitz (2009). The experiment was conducted in a dark experimental room. Participants were instructed to fixate a 0.2° red fixation point presented at the center of the screen and to avoid tracking any particular dots during the stimulus presentation. In each trial, participants were presented with the RDM for a Duration of either 200 or 800 ms and then had 4 s to report the perceived direction of motion by clicking on one of four bars that corresponded to the possible directions of motion (Figure 1). If participants' response was within 22.5° of the presented direction, then participants received visual feedback that their response was correct. For every trial, participants were instructed to fixate on the central spot.

All participants completed two experimental sessions, one for each Platform, in counterbalanced order. In each session, participants first completed a practice block consisting of 10 trials to familiarize themselves with the procedure. The method of constant stimuli was employed, such that the random dot motion was presented at two fixed Durations (200 or 800 ms) at 10 different Coherence levels (2, 4, 6, 8, 10, 15, 20, 25, 30, and 50%) and at one of four directions (45, 135, 225, and 315 degrees). Each session was comprised of 640 trials that were divided into four task blocks, with a self-paced break between each block. The RDM paradigm used in this study is available online on Inquisit Millisecond's library.¹

Data analysis

Percent correct as a function of coherence is calculated as the total number of correct trials (per coherence level) divided by the total number of trials (per coherence level). Based on this an accuracy value between 0 and 1 is obtained and multiplied by 100 for reporting percent correct. To characterize performance, a three-way repeated measures ANOVA was run using the statistical platform JASP (JASP Team (2022). JASP (Version 0.16.3) [Computer software]), with Coherence, Duration, and Platform as the repeated measures factors, and the order of platform as the between subject factors. We used the same statistical software,

¹ <https://www.millisecond.com/download/library/rdk>

JASP, to conduct Pearson's correlation analysis between the average performance score for each subject on MATLAB versus Milliseconds.

Results

Overall, both MATLAB and Inquisit Millisecond Platforms led to well-characterized psychometric functions that were quite similar between the Platforms (Figure 2, RTs can be found in Supplementary Figure S1). To characterize performance, we performed a three-way repeated-measures ANOVA that evaluated the effects of Coherence, Platform (Inquisit or MATLAB), and Duration on participants' behavioral performance. As expected, we found a significant effect of Coherence [$F(9, 378) = 139.0$, $p < 0.001$, $\eta_p^2 = 0.768$], and Duration [$F(1, 42) = 78.0$, $p < 0.001$, $\eta_p^2 = 0.651$]. Importantly, we found no significant effect of Platform [$F(1, 42) = 0.755$, $p < 0.390$, $\eta_p^2 = 0.018$]. However, there was a significant interaction between Coherence and Platform [$F(9, 378) = 6.0$, $p < 0.001$, $\eta_p^2 = 0.125$] and between Coherence and Duration [$F(9, 378) = 10.4$, $p < 0.001$, $\eta_p^2 = 0.197$]. Post-hoc comparisons can be found in Supplementary Tables S1, S2 in the Supplementary Figures. Overall, these results show that participants' performance increases as a function of Coherence ($\eta_p^2 = 0.768$) and Duration ($\eta_p^2 = 0.651$), and that these relationships are not significantly affected by the Platform ($\eta_p^2 = 0.125$). These data suggest that while there are some differences between the psychometric functions – for example, a slight advantage for Inquisit at the brief viewing Duration – there is substantial similarity in performance across the two platforms.

A key consideration in evaluating the relationship between the platforms is the extent to which they are evaluating the same construct. To address this, we examined inter-platform reliability.

This was addressed by correlating the mean of each participant's performance between platforms (Figure 3). We found that there were high inter-platform correlations for both viewing Durations (200 ms, $r = 0.866$, $p < 0.001$ and 800 ms, $r = 0.870$, $p < 0.001$). The significant inter-platform reliability shows that the Inquisit Millisecond platform leads to measurements that are largely consistent with RDM displays generated by MATLAB with Psychtoolbox. For context, we also examined intra-test correlations. To estimate these, we averaged performance accuracy across the first and fourth experimental runs and analyzed its correlation with the average of the second and third experimental runs. These intra-platform correlations (Figure 4) were also quite high for MATLAB with Psychtoolbox (200 ms, $r = 0.90$, $p < 0.001$ and 800 ms, $r = 0.96$, $p < 0.001$) and Inquisit (200 ms, $r = 0.97$, $p < 0.001$, and 800 ms, $r = 0.94$, $p < 0.001$). We note that while intra-platform correlations are numerically greater than inter-platform correlations, the intra-platform correlations are across sessions, while inter-platform correlations are within sessions. Overall, these data suggest that both platforms produce reliable estimates of motion perception, and that inter-platform reliability is reasonably comparable to intra-platform reliability.

Discussion

The main contribution of this work is the validation of a cross-platform novel RDM method that can be run on participants' own devices without the need for MATLAB and Psychtoolbox and is easily accessible to researchers, students, and clinicians. To validate a generic RDM-based task on the Inquisit Millisecond platform, in this study, we directly compared implementation and results from the same RDM task in MATLAB *via* Psychtoolbox and Inquisit Millisecond. Results

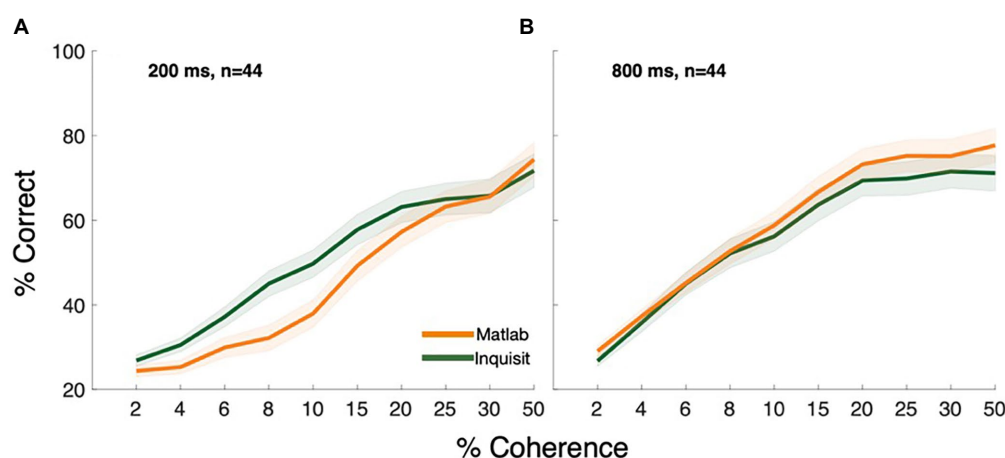


FIGURE 2
Coherence response functions of MATLAB and Inquisit Millisecond at viewing Durations (A) 200 ms and (B) 800 ms. Error is plotted as the standard error of the mean across all participants.

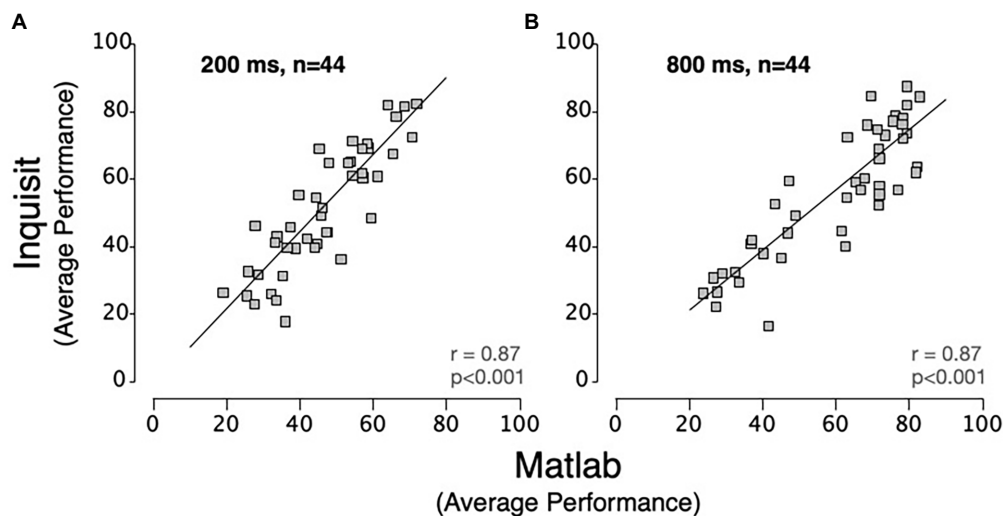


FIGURE 3
Scatter plots showing inter-platform relationships between MATLAB and Inquisit Millisecond for (A) 200 ms and (B) 800 ms RDM Durations. Each data point represents the mean performance of a single participant for each platform. The lines are trendlines fitted to the data (Pearson's correlation: $p < 0.001$ for both 200 ms and 800 ms duration).

show that psychometric functions of performance as a function of coherence are largely the same between MATLAB with Psychtoolbox and Inquisit Millisecond. Further, we found robust intra-platform and cross-platform reliabilities. These findings demonstrate that Millisecond RDM psychometric properties are consistent with MATLAB/Psychtoolbox and that the Millisecond RDM can provide high-quality and reliable data.

We note that the creation and presentation of RDM stimuli typically require technical sophistication and that the contribution of an accessible RDM that can be run remotely is particularly valuable given the increased focus on remote data collection that has been spurred by the COVID pandemic. While other platforms, including Psychophysics Toolbox for MATLAB and PsychoPy, can also support remotely-administered studies on participants' own devices, these platforms require greater technical skills to use than the Millisecond Platform (and, in the case of MATLAB/Psychtoolbox, may involve installing commercial software that is not free). Given prior results showing that performance on RDM tasks is highly sensitive to the exact parameters used (Pilly and Seitz, 2009), it is particularly valuable to have an accessible RDM stimulus generator that has known psychometric properties.

However, we also note that there are some important limitations to the Millisecond RDM stimuli. The first issue is that while we compared performance between MATLAB and Millisecond on the same computer and the same monitor, the extent to which the Millisecond RDM performs consistently across different platforms remains to be clarified, as displays cannot be as easily controlled when using participant-owned devices. Of note, we did not address the extent to which frames may have been dropped or the extent to which this may be different between the platforms when run on different displays.

However, we note that this is a limitation of any RDM that would be run on a participant's device. We also note that while generally, in vision science, parameters are coded as degrees per visual angle, in Millisecond Inquisit, parameters are coded as percent screen size. However, given that visual angle is dependent upon the distance of the participant from the screen and percent screen size is dependent on the screen size, both Inquisit Millisecond and MATLAB with Psychtoolbox have limitations in uncontrolled settings, and both can easily be matched when the screen size is known. Additionally, online remedies such as virtual chinrests (Li et al., 2020) can be employed to minimize the change in stimulus perception. We note that for studies dependent upon a single measurement, these issues can be a significant concern; however, studies involving within-person designs, such as interventions or repeated testing, may be less impacted by these limitations.

Conclusion

Overall results support using Millisecond RDM as a reliable method for characterizing motion discrimination performance in healthy human subjects. While further research will be required to understand the extent to which the Millisecond RDM is reliable when run at home on participant's own devices, the present research sets the stage for this next step in validation. We do note that for any remotely-administered study, whether using Millisecond or another platform, careful design of the experimental parameters and cross-checking units may be required to ensure that data will be comparable to previously-reported results. Still, even with these limitations, this new RDM stimulus creation and deployment software can be of benefit to

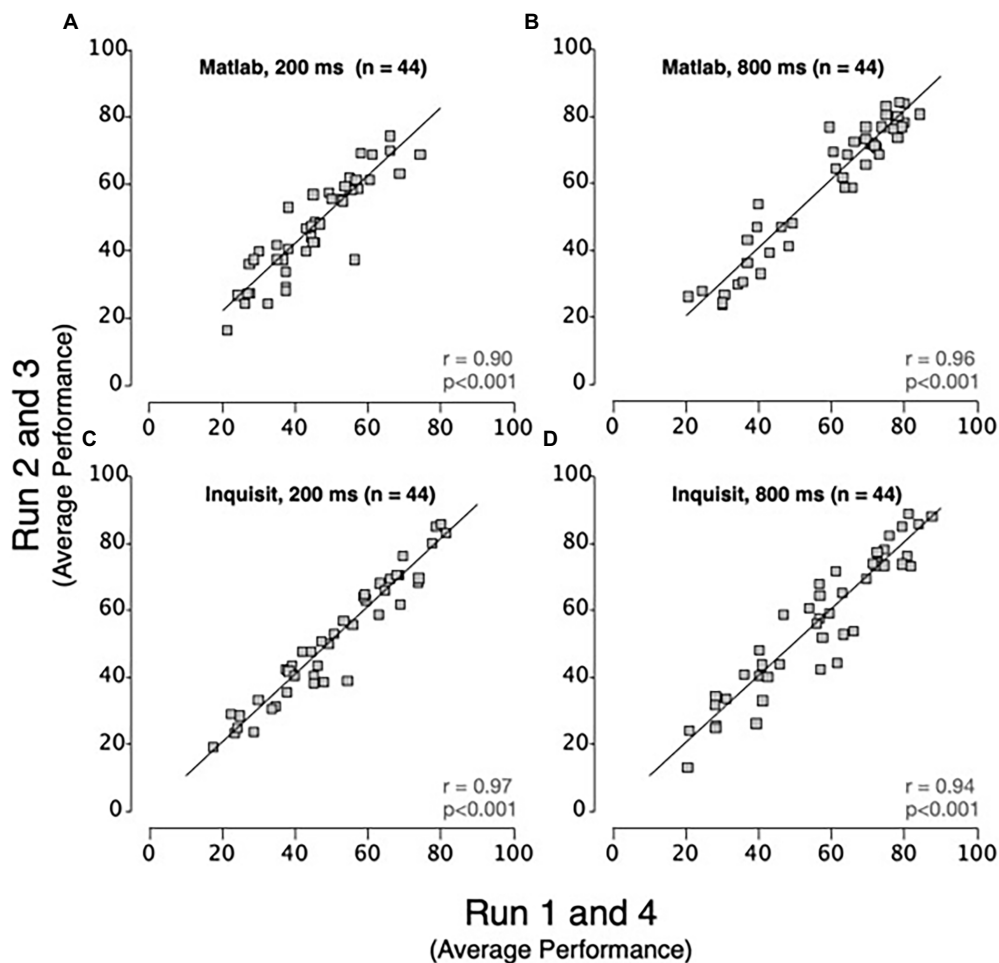


FIGURE 4

Scatter plots showing intra-platform relationships for MATLAB and Inquisit Millisecond platforms at 200 ms and 800 ms RDM durations. Each data point represents the mean performance of a single participant for runs 1 and 4 vs. runs 2 and 3 on MATLAB at (A) 200 ms and (B) 800 ms and on Inquisit Millisecond at (C) 200 ms and (D) 800 ms. The line is a trendline fitted to the data (Pearson's correlation: $p < 0.001$ for both 200 ms and 800 ms durations on MATLAB and Inquisit Millisecond).

researchers, students, and clinicians as a reliable method to understand human motion processes.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors upon request, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by University of California, Riverside (UCR) Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

Author contributions

RDM algorithm in MATLAB was written by AS. Millisecond Inquisit RDM was written by Katja Borchert (Senior Consultant at Millisecond Software LLC) in consultation with AS and KY. Data were collected by undergraduate students SK, SA, and HK. Data analyses were conducted by KY. All authors assisted in the writing of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was partially supported by a Canadian Institute for Advanced Research Azrieli Global Scholars Fellowship (to MP). The funders had no role in the study design, analysis, or conclusion.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2022.1035518/full#supplementary-material>

References

- Adelson, E. H., and Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A* 2:284. doi: 10.1364/josaa.2.000284
- Ball, K., and Sekuler, R. (1982). A specific and enduring improvement in visual motion discrimination. *Science* 218, 697–698. doi: 10.1126/science.7134968
- Banton, T., Dobkins, K., and Bertenthal, B. I. (2001). Infant direction discrimination thresholds. *Vis. Res.* 41, 1049–1056. doi: 10.1016/s0042-6989(01)00027-x
- Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., et al. (2008). Probabilistic population codes for Bayesian decision making. *Neuron* 60, 1142–1152. doi: 10.1016/j.neuron.2008.09.021
- Benton, C. P., and Curran, W. (2009). The dependence of perceived speed upon signal intensity [review of the dependence of perceived speed upon signal intensity]. *Vis. Res.* 49, 284–286. doi: 10.1016/j.visres.2008.10.017
- Bianchi, R., and Laurent, E. (2015). Emotional information processing in depression and burnout: an eye-tracking study. *Eur. Arch. Psychiatry Clin. Neurosci.* 265, 27–34. doi: 10.1007/s00406-014-0549-x
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat. Vis.* 10, 433–436.
- Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S., and Movshon, J. A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis. Neurosci.* 13, 87–100. doi: 10.1017/s095252380000715x
- Britten, K. H., Shadlen, M. N., Newsome, W. T., and Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci. Off. J. Soc. Neurosci.* 12, 4745–4765. doi: 10.1523/JNEUROSCI.12-12-04745.1992
- Cleary, R., and Braddick, O. J. (1990). Direction discrimination for band-pass filtered random dot kinematograms. *Vis. Res.* 30, 303–316. doi: 10.1016/0042-6989(90)90045-m
- Gold, J. I., and Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.* 30, 535–574. doi: 10.1146/annurev.neuro.29.051605.113038
- Hillman, C. H., Rosengren, K. S., and Smith, D. P. (2004). Emotion and motivated behavior: postural adjustments to affective picture viewing. *Biol. Psychol.* 66, 51–62. doi: 10.1016/j.biopsycho.2003.07.005
- Jin, Y.-R., and Lin, L.-Y. (2022). Relationship between touchscreen tablet usage time and attention performance in young children. *J. Res. Technol. Educ.* 54, 317–326. doi: 10.1080/15391523.2021.1891995
- Kim, H. R., Angelaki, D. E., and DeAngelis, G. C. (2016). The neural basis of depth perception from motion parallax. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 371:20150256. doi: 10.1098/rstb.2015.0256
- Kim, Y.-G., and Mollon, J. D. (2002). Conditions under which stereopsis and motion perception are blind. *Perception* 31, 65–71. doi: 10.1068/p3209
- Kleiner, M., Brainard, D., and Pelli, D. (2007). “What’s new in Psychtoolbox-3?” *Perception* 36 ECVF Abstract Supplement.
- Law, C.-T., and Gold, J. I. (2008). Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. *Nat. Neurosci.* 11, 505–513. doi: 10.1038/nn2070
- Li, Q., Joo, S. J., Yeatman, J. D., and Reinecke, K. (2020). Controlling for participants’ viewing distance in large-scale, psychophysical online experiments using a virtual chinrest. *Sci. Rep.* 10:904. doi: 10.1038/s41598-019-57204-1
- Mazurek, M. E., Roitman, J. D., Ditterich, J., and Shadlen, M. N. (2003). A role for neural integrators in perceptual decision making. *Cereb. Cortex* 13, 1257–1269. doi: 10.1093/cercor/bhg097
- Nadler, J. W., Angelaki, D. E., and DeAngelis, G. C. (2008). A neural representation of depth from motion parallax in macaque visual cortex. *Nature* 452, 642–645. doi: 10.1038/nature06814
- Pavan, A., Boyce, M., and Ghin, F. (2016). Action video games improve direction discrimination of Parafoveal translational global motion but not reaction times. *Perception* 45, 1193–1202. doi: 10.1177/0301006616663215
- Pavan, A., Ghin, F., and Campana, G. (2021). Visual short-term memory for coherent and sequential motion: a rTMS investigation. *Brain Sci.* 11:471. doi: 10.3390/brainsci11111471
- Pavan, A., Langgartner, D., and Greenlee, M. W. (2013). Visual short-term memory for global motion revealed by directional and speed-tuned masking. *Neuropsychologia* 51, 809–817. doi: 10.1016/j.neuropsychologia.2013.02.010
- Pilly, P. K., and Seitz, A. R. (2009). What a difference a parameter makes: a psychophysical comparison of random dot motion algorithms. *Vis. Res.* 49, 1599–1612. doi: 10.1016/j.visres.2009.03.019
- Pelli, D. G. (1997). The video toolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442.
- Purushothaman, G., and Bradley, D. C. (2005). Neural population code for fine perceptual decisions in area MT. *Nat. Neurosci.* 8, 99–106. doi: 10.1038/nn1373
- Seitz, A. R., Naney, J. E., Holloway, S. R., Koyama, S., and Watanabe, T. (2005). Seeing what isn’t there; the costs of perceptual learning. *Int. J. Vis.* 5:864. doi: 10.1167/5.8.864
- Seitz, A. R., and Watanabe, T. (2003). Is subliminal learning really passive? *Nature* 422:36. doi: 10.1038/422036a
- Smith, D. P., Hillman, C. H., and Duley, A. R. (2005). Influences of age on emotional reactivity during picture processing. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 60, P49–P56. doi: 10.1093/geronb/60.1.p49
- Smolker, H. R., Wang, K., Luciana, M., Bjork, J. M., Gonzalez, R., Barch, D. M., et al. (2022). The emotional word-emotional face Stroop task in the ABCD study: psychometric validation and associations with measures of cognition and psychopathology. *Dev. Cogn. Neurosci.* 53:101054. doi: 10.1016/j.dcn.2021.101054
- Stewart, N., Chandler, J., and Paolacci, G. (2017). Crowdsourcing samples in cognitive science. *Trends Cogn. Sci.* 21, 736–748. doi: 10.1016/j.tics.2017.06.007
- Van Damme, S., Crombez, G., and Lorenz, J. (2007). Pain draws visual attention to its location: experimental evidence for a threat-related bias. *J. Pain: Official J. American Pain Society* 8, 976–982. doi: 10.1016/j.jpain.2007.07.005
- Van Damme, S., Lorenz, J., Eccleston, C., Koster, E. H. W., De Clercq, A., and Crombez, G. (2004). Fear-conditioned cues of impending pain facilitate attentional engagement. *Neurophysiologie Clinique = Clin. Neurophysiol.* 34, 33–39. doi: 10.1016/j.neucli.2003.11.001
- Vogt, J., De Houwer, J., Koster, E. H. W., Van Damme, S., and Crombez, G. (2008). Allocation of spatial attention to emotional stimuli depends upon arousal and not valence. *Emotion* 8, 880–885. doi: 10.1037/a0013981
- Watanianuk, S. N., and Sekuler, R. (1992). Temporal and spatial integration in dynamic random-dot stimuli. *Vis. Res.* 32, 2341–2347. doi: 10.1016/0042-6989(92)90097-3
- Zakrzewska, M. Z., and Brzezicka, A. (2014). Working memory capacity as a moderator of load-related frontal midline theta variability in Sternberg task. *Front. Hum. Neurosci.* 8:399. doi: 10.3389/fnhum.2014.00399
- Zanker, J. M. (1999). Perceptual learning in primary and secondary motion vision. *Vis. Res.* 39, 1293–1304. doi: 10.1016/s0042-6989(98)00234-x
- Zohary, E., Celebrini, S., Britten, K. H., and Newsome, W. T. (1994). Neuronal plasticity that underlies improvement in perceptual performance. *Science* 263, 1289–1292. doi: 10.1126/science.8122114



OPEN ACCESS

EDITED BY

Alyssa A. Brewer,
University of California, Irvine,
United States

REVIEWED BY

Gloria Tachie-Donkor,
University of Cape Coast, Ghana
Christopher Owusu-Ansah,
Akenten Appiah-Menka
University of Skills Training and
Entrepreneurial Development, Ghana

*CORRESPONDENCE

Wenyan Yu
ywy@zjut.edu.cn
Yanqi Wu
wuyanqi@zjut.edu.cn

SPECIALTY SECTION

This article was submitted to Perception
Science, a section of the journal Frontiers
in Psychology

RECEIVED 05 July 2022

ACCEPTED 23 November 2022

PUBLISHED 04 January 2023

CITATION

Yu W, Jiang Y, Wu Y and Cheng Y (2022)
Multisensory reading promotion in
academic libraries.
Front. Psychol. 13:987180.
doi: 10.3389/fpsyg.2022.987180

COPYRIGHT

© 2022 Yu, Jiang, Wu and Cheng. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that
the original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Multisensory reading promotion in academic libraries

Wenyan Yu*, Yiping Jiang, Yanqi Wu* and Yanxia Cheng

Library, Zhejiang University of Technology, Hang Zhou, Zhejiang, China

To confront college students' new reading patterns and the continuous decline in academic library borrowing rates, we conducted empirical research on promoting multisensory reading as a way to attract students' attention, and to stimulate interest in, and promote the practice of, reading through a library program called "Reading Today Listening Everyday" (RTLE) on a library's WeChat public account. The program involved 48 librarians and 105 students who were recruited into different groups to co-create, edit and release multisensory tweets every workday. Multisensory contents including text-based content, audio-based content and emotional resonance were presented to evoke readers' visual, audio, and emotional senses to induce more reading practice. Using the Context, Input, Process and Product (CIPP) evaluation method, the multisensory presentation in RTLE program was proven to be effective in promoting library reading with a high number of tweeted page views and an increased borrowing rate for recommended books. In 2020, 269 issues accompanied by 269 audio frequencies garnered 80,268 page views, depending on the caliber of the reading promoter out of the 48 librarians and 52 student anchors behind it. The 484 RTLE-recommended books were borrowed 113 times in 2020, which was a rate 1.46 times higher than in 2019 (77 times). The analysis of the relationship between tweet views and borrowing rates for recommended books indicates that more page views indicate greater reader interest, leading to increased borrowing. From readers' feedback and comments, the gain afforded by multisensory reading can improve higher-level reading trends such as the number of reading interests, enjoyment, engagement, etc.

KEYWORDS

multisensory reading, reading promotion, visual, audio, emotional, academic library

Introduction

The notion that the senses are better conceptualized as interrelated modalities rather than independent channels is supported by many studies, providing evidence for common neural and psychological mechanisms for the processing of multisensory information (Spence et al., 2000; Doyle and Walker, 2002). For example, for a public visiting a museum, simply looking at art works may not be sufficient to interpret them. Sensitivity to additional cues, like background music and audio commentaries, will also guide them to apprehend them. Similarly, one may consider the multisensory attributes of reading. In the process of

reading, multisensory reading which involves more readers' senses should be more interesting and effective. Considering the changing in mass reading habits and strengths of multisensory presentation, the library of Zhejiang University of Technology tries to introduce multisensory mode into reading promotional activities to recommend library collections.

When analyzing changes in college students' reading patterns, a "sensory turn" could be identified. According to the report of the 18th National Reading Survey in 2021 (*The Eighteenth National Reading Survey Report, 2021*), the comprehensive reading rate on social media continues to grow steadily, for example, content reading rate on mobile phone for adults was 79.4%, and content hearing rate was 31.6%. Compared with traditional reading, digital reading integrates text, sound, image, and voice, which can bring readers a multisensory and tridimensional reading experience. Intriguingly, and contrary to the increasing reading rate, the book-borrowing rate of many renowned university libraries fell over 50% in recent years (Zheng, 2019), which calls for research on how librarians conduct reading promotional activities to meet students' growing reading needs and promote the utilization of library collection.

Despite the heartening proliferation of promotional reading programs in academic libraries, many studies reported only their limited success and sustainability (Gauder et al., 2007; Bosman et al., 2008; Fajardo, 2010; Chesnut, 2011; Mackay and Tarulli, 2015; Dali and McNiff, 2020). We view the lack of a differentiated approach to promote reading and the low level of reader engagement in activities as one of the reasons for the limited success and short lifespan of reading initiatives. Unfortunately, there are unexpectedly few studies on this subject. To enrich the literature, We conducted empirical research on reading promotion of multisensory mode as an attempt to attract students' attention, stimulate reading interest, and promote reading practice, which was implemented in a library program called "Reading Today Listening Everyday" (RTLE) on a library's WeChat public account. We aimed for our results to provide a new mode to improve library reading promotion with the expectation to meet readers' needs and increase the utilization rate of library collection. Additionally, the experience of multisensory reading promotion allows librarians to innovate when providing reading services and provide a reference for reading promotion for other libraries, businesses, and policy-planning offices.

Literature review

While the multisensory aspects of handling books mainly has to do with perception, the multisensory aspects of reading mostly involve the rich mental imagery that is evoked by various forms of presentation (Spence, 2020). Mental imagery is often described as "seeing with the mind's eyes" and "listening with the mind's ears," and it plays an important role in understanding individual cognitive function (Kosslyn, 2005). The multiple sensory aspects of mental imagery that may be at work during reading mainly

include visual imagery (Boerma et al., 2016; Brosch, 2018), auditory imagery (Perrone-Bertolotti et al., 2012; Moore and Schwitzgebel, 2018), and the complementary "inner ear" non-speech imagined sounds that may be evoked in a passage (Brunyé et al., 2010). There is even some evidence that readers may give different characters different voices too (Alexander and Nygaard, 2008; Kurby et al., 2009). Mental imagery may also be triggered in other sensory modalities while reading like olfactory and somatosensory mental imagery. When reading olfactory descriptors, or words related to a distinctive smell, such as the word cinnamon, has been shown to give rise to increased activation in olfactory brain areas (González et al., 2006). What is more, to the extent that the digital reading enhancing multisensory aspects of reading might provide opportunity to engage the reader's mental imagery capacities so fully. Crucially, multisensory presentation in evoked mental imagery supports readers in the learning process, from language comprehension (Bottini et al., 1994), to socially-motivated behaviors such as perspective taking (Decety and Ruby, 2001), to motor learning (Yägüez et al., 1998). Researchers have reported multisensory approach was not only effective for students with dyslexia or reading difficulties who are learning to read and write (Warnick and Caldarella, 2016; Soliman and Al-Madani, 2017; Gori et al., 2020), but could also facilitate student learning outcomes, such as the integration between vocal signals and visual cues (like facial movements) in social and language learning (King and Lewkowicz, 2012; Weatherhead and White, 2017), the association of spoken words and visually presented objects in the context of vocabulary learning (Pereira et al., 2014), and the acquisition of speech sound-letter correspondences during reading development (Horbach et al., 2015). In recent years, multisensory experience was also designed for library reading promotion, book exhibition, music reading and art appreciation (Spence, 2020; Arora, 2021; Paraskevopoulos et al., 2021).

Recognizing the importance of multisensory attributes, institutions have begun to organize events. In 2013, the Canadian Library participated in a reading promotion event known as "Words on the Street," which featured book exhibitions, handicrafts, painting, role playing and other multisensory activities and attracted the attendance and participation of 270,000 people (Yan, 2016). In Oxford's Bodleian Library (2022) conservators often note that something from the Duke Humfrey's Library has been brought in, based on nothing more than the smell that pervades the air when an item from the historic old reading room arrives. In order to help people cope with pandemic lockdown, the Bodleian Library released a stream of sounds normally heard in university libraries such as creaks, rustling, coughs and traffic noises, even including a recording from the Duke Humfrey's Library (Kidd, 2020). Recognizing the importance of multisensory attributes, sensory events have been organized at St. Paul's Cathedral Library (Bembibre and Strlič, 2017), the Birmingham Museum and Art Gallery, and, in late 2020, the book exhibition at the Weston Library, part of Oxford's Bodleian Library (De Bruxelles, 2020). In China, "Modeling

Library” activities in Xiamen Library in 2020 provided multisensory experiences, which made children’s reading more vivid and interesting (Song, 2022). By analyzing activities at the “Chinese Library Reading Club,” Zhongshan Library of Guangdong Province in 2018 concluded that public libraries also needed to conduct multisensory reading promotional activities for minors to effectively promote the quality and efficiency of services (Song, 2022). For academic libraries, multisensory reading promotion has been proposed as an innovative form suitable to various reader groups (Zhang and Zang, 2018). The library of Zhengzhou University in 2021 set up the “Book Sounds in Zhengzhou University” column by collecting students’ audio literary works read aloud by students, accompanied by the text and illustrations of the books, allowing readers to experience the mood and connotation of the literary works from both the visual and auditory senses (Song, 2022). Liu designed a multisensory model for reading promotion in academic libraries in 2019, showing that the mode could motivate college students’ reading, provide a multisensory reading experience, and ultimately achieve the purpose of independent and deep reading (Liu, 2019).

Many scholars have come to acknowledge that multisensory presentation [for instance, how a curator might attempt to recreate the sights, sounds, and smells associated with the Canterbury Tales without it necessarily relating to the book itself (Aggleton and Waskett, 1999)] is an innovative and effective mode for promoting reading in libraries. Yet, based on the above literature, there is little research literature on multisensory reading promotion in academic libraries and long-term extensive research is insufficient. By identifying the research gaps, we derive the following research questions to be explored: how can multisensory presentations be integrated into library reading promotional activities, and how effective is multisensory reading promotion in improving utilization of library collections? Matusz et al., (2019) pointed out that multisensory contexts reflecting naturalistic settings, provide an adaptive benefit for learning. Logically, the gain afforded by multisensory presentation can improve higher-level reading trends such as the number of reading interests, level of enjoyment, engagement etc. Therefore, we conducted an experimental study for more than 1 year on multisensory mode reading promotion implemented in the RTLE program. We expect the study results could answer the above research questions and provide a reference for other libraries.

Methodology

Our research used the case study approach to evaluate the effectiveness of the RTLE reading promotion program, which is carried out on a library WeChat public account. WeChat is the most popular mobile application among college students and has already been employed in 84.6% of Chinese well-known university libraries (Wei and Yang, 2017). The project involved 48 librarians and 105 students who were recruited into different groups to co-create, edit, and release tweets every workday. To achieve

eye-catching, ear-occupying, and heart-touching effects, RTLE presents multisensory content and encourages readers’ interactions on the platform to evoke their visual, audio, and emotional senses. RTLE effectiveness in reading promotion was also determined using the Context, Input, Process and Product (CIPP) evaluation method to measure context, input, process, and results. In the section that follows, we take a closer look at multisensory presentation in the RTLE program and the effectiveness of multisensory reading promotion on the utilization of library collections.

Multisensory reading contents

Text-based content

Traditionally, librarians or students organize and participate in text writing unilaterally. This is in contrast with the RDLE program, whose contents are jointly created by them. All librarians participate in creating essays to highlight the contents of good books, while the written part of tweet could also be produced by the excellent students recruited from activities of the award solicit article. To increase reading interest and attention, the text part of each tweet is blended with theme-relevant pictures, characters’ portraits, and a story background introduction. Before their release, the program committee carries out a final polishing of the texts to match them with the selected theme. Cover pictures, library codes, and library location are included in each tweet, so that students can promptly find the recommended books.

Audio-based content

The program, committed to transmitting the beauty of audio reading and enhancing communication with readers, has invited students majoring or interested in broadcasting to obtain sound recordings. Approximately 3–5 min of audio frequency as reading guidance were generated by students from the excerpts of the recommended books. The anchor’s introduction is also present at the beginning of each tweet to encourage other students to take part in the broadcasting work. By listening to the classic excerpts, readers vividly comprehend the spirit of the book, and its breadth and depth can be extended.

Emotional resonance

It is difficult to achieve an emotional relationship with a text in the absence of participation. In RTLE, librarians’ and students’ engagement generates both service and products such as text and audio frequency. They make use of their own knowledge and give full play to their own expertise on understanding the recommended books in deeply collation and refinement from a unique personal perspective. Relying on the social attributes of the WeChat platform, readers’ feedback and comments on contents or program have been encouraged and collected. In that way, readers’ sympathetic responses echo to the mood that refers to a state of creators’ understanding of the objective.

Effectiveness of multisensory reading

Based on the CIPP evaluation method, originally proposed by L.D. Stufflebeam, a famous American evaluation expert (Lv, 2021), the effectiveness of RTLE multisensory reading promotion is analyzed to optimize services through a continuous “evaluation-feedback” process. Context evaluation focuses on readers’ needs, existing problems, and opportunities before the program is conducted. Input evaluation, based on the context evaluation, analyzes the required resources, conditions, and feasibility of the program. Process evaluation collects data and provides feedback throughout program implementation. Result evaluation makes use of quantitative or qualitative methods to examine the degree of accomplishment of the expected objectives.

According to the CIPP evaluation method together with the RTLE reading promotion process, the evaluation steps were as follows: Context evaluation considered readers’ multisensory and fragmented reading needs and the dilemma of the continuous decline of book lending in academic libraries (mentioned in the Introduction). Input evaluation analyzed human resources, the WeChat platform, the number and types of recommended books and the multisensory presentation involving original text, audio frequency and pictures. Process evaluation collected page views and re-postings of tweets along with readers’ feedback. Result evaluation calculated the borrowing volume of recommended books. Since its launch, the RTLE program has made great strides and established a good reputation, capturing positive comments from readers. The quantitative and qualitative evaluations of the RTLE program are analyzed in the sections that follow.

Reading today listening everyday reading promotion data

The achievements of RTLE of multisensory reading promotion throughout 2020 are described in Table 1. All its tweets have been classified by theme into four categories: literary, cultural, historical, and comprehensive. Comprehensive themes include technology, stories of role models, and anti-coronavirus knowledge. The number and page views of different themes are shown in Table 2.

One tweet per weekday is released on the library public WeChat account of Zhejiang University of Technology, and the weekend release is added when it comes to special days such as traditional festivals, anniversary of famous people etc. In 2020, 269 issues were released, with 80,268 page views. Page view growth is largely a function of the appeal of the content of RDLE program, which depends on the caliber of the reading promoter behind it. In RTLE, 52 student anchors and 48 librarians of different backgrounds have exploited their advantages so that every issue was told vividly and freshly from a particular viewpoint. Students

participated in recording 269 audios and interacted with their comments, while librarians wrote over 800,000 words of original text and recommended 484 classic books for students. Although students are easily attracted by “fast-food type” entertainment culture, we are seeing more willingness to participate and interact in the program, with 2,108 active re-postings.

As shown in Table 2, page views varied by theme. Consistent with the statistical results, we found that students were most interested in the themes of technology, role models, and anti-coronavirus knowledge, with an average of 321.37 page views. The factor of psychological needs could be, at least in part, at play. The coronavirus pandemic, initiated at the start of 2020, aroused students’ anxiety and stress due to uncertainties in its control. The multisensory mode reading promotion attracts students’ attention and the gain of reading from relative scientific approaches, anti-coronavirus knowledge and books helps build up students’ confidence to fight against the pandemic, so that the sort of theme was most welcomed. Additionally, another highly performing theme was the literary one, with 300.21 average page views, which conveyed the spirit of classic works such as poems, prose, and novels. Tweets with cultural and historical themes, respectively earning 290.15 and 283.34 average page views, have appealed well to students too.

The daily tweets were previously described in an institutional blog created by the library on the Colmenarejo campus of University Carlos III of Madrid with the goal of promoting reading among the university population (Hernandez and de la Cruz, 2007). At present, most academic libraries measure the effectiveness of reading promotion activities by counting the number of participants, the scale of activities, the number of books borrowed during and after the activities, and readers’ feedback (Lan, 2021). In 2020, 78 book-recommendation tweets were released on the library WeChat public account of Shenzhen University of Technology, with page views of 11,325 in total and 145 on average (Lv, 2021). Ma has investigated six Fine Arts Colleges and found that the book-recommendation tweets on their library WeChat public accounts have been read 6,312 times in total and 117.22 on average (Ma, 2018). By pouring more input into this multisensory and interesting mode, the RTLE program has made great strides in terms of increasing the number of page views and tweets.

Data of borrowing volume

To understand the relationship between tweet views and the borrowing rate of the recommended books, page views of 269 tweets were divided into 4 ranks based on their number of views: more than 1,000 which is denoted by 1,000↑, 500 -1,000, 300 -500, and less than 300 which is denoted by ↓300. Furthermore, the

TABLE 1 The achievements of RTLE program.

Issues	Audio frequencies	Recommended books	Participating librarians	Student anchors	Page views	Re-posting
269	269	484	48	52	80,268	2,108

borrowing rates in 2020 and 2019 were compared and analyzed to examine the effect of multisensory reading promotion.

Although the borrowing rate of paper books has been significantly reduced in current times, the borrowing rate of RTLE-recommended books has increased surprisingly. As shown in Table 3, the books recommended in tweets with more than 1,000 and 500–1,000 page views have been borrowed 19 and 27 times, respectively. Both of them are over 1.69 times higher than those of the year before. The 145 books recommended in tweets with 300–500 page views also had a good performance, being borrowed 34 times, which is 1.48 times higher than their borrowing rate in the previous year. The 181 books recommended in tweets with less than 300 page views were borrowed 33 times, which is just a bit higher than their borrowing rate in the previous year. Overall, the 484 recommended books were borrowed 113 times in 2020, which is 1.46 times higher than this value in 2019 (77 times). Although the sample size of this study is small and limited to one library, we also can get some implications. Based on statistical results, more page views indicate more reader interest leading to increased borrowing. Reading in a multisensory mode is helpful for book borrowing, especially in terms of recommended books with higher page views.

Book recommendations in the RTLE program continued with the launch of the “RTLE @Readers’ Talk” column in 2021. In 2020, the pictorial and audio story was used as a core reading guideline to recommend books in the “RTLE @ Librarians’ Writing” column, while in 2021, interpretation of the recommended books was regarded as the key point to encourage reader feedback in the “RTLE @ Readers’ Talk” column. Using the recommendation list as the outline and book interpretation as the core, “RTLE @ Readers’ Talk” initiated teachers and students on writing book

reviews. Using the RIA note reading method (original reading + concise restatement + relative application), readers concentrated and simplified the content, extracted its essence, and wrote tweets to publish on the library WeChat public account for every issue. As of July 31, 2021, 140 original audio book reviews have been published by readers. Meanwhile, in June 2021, based on a new weekly book bulletin by Yun Reading Platform, “RTLE @ To Know New Books” column, accompanied by excellent audio fragments, was launched to recommend 4–5 selected new books to readers.

In addition, the library set up customized bookshelves to display RTLE-recommended books. These RTLE shelves, different from the traditional fixed and somber counterparts, capture readers’ attention by their various shapes and vibrant red, bright yellow and ocean blue colors. The recommended books have also been put on shelves and exhibited regularly by librarians. Meanwhile, the borrowed books have been counted to analyze the effects of reading promotion. For instance, not included in Tables 3, 10 of 17 RTLE-recommended books were borrowed during the 2020 Reading Festival after 2 weeks of first-time exhibition on bookshelves. Their borrowing rate surprisingly reached 58.8%. This was not a singular case: other libraries also reported enhanced lending rate after promotional reading activities. For instance, after holding the book reading and sharing club “Reading Joy,” the relative lending rate of three recommended books at the Shenzhen University of Technology Library increased 184, 200 and 133%, respectively (Lv, 2021). Lin used the “360°-pleasant reading” carried out by the Hainan Medical College Library as a practical example; through in-depth guided reading, the lending rate of recommended books increased by 177.4, 189.2, and 170.8%, respectively (Lin, 2015). By highlighting recommendations and using a centralized display, several inactive books in the library have been reintroduced into circulation.

Readers’ feedback

Since its launch, RTLE has attracted the attention of more than 40,000 readers who often leave positive messages in the comment area to express their satisfaction. According to messages in the comment area, reader feedback can be summarized into the categories shown in Table 4.

From the readers’ feedback, RTLE generally impressed readers. Readers’ messages expressed their love for the column

TABLE 2 Number and page views of different themes.

Theme	Number of tweets	Page views	Average page views of each tweet
Comprehensiveness	62	19,925	321.37
Literature	74	22,216	300.21
Culture	65	18,860	290.15
History	68	19,267	283.34
Total	269	80,268	298.39

TABLE 3 The borrowing volume of the recommended books.

Page views	Number of the recommended books	The borrowing volume of the recommended books in 2020	The borrowing volume of the recommended books in 2019	Borrowing rate after and before recommendation
1,000↑	61	19	11	173%
500–1,000	97	27	16	169%
300–500	145	34	23	148%
↓300	181	33	27	122%
Total	484	113	77	146%

TABLE 4 Readers' feedback.

	Category	Messages
1	Reading desire has been stimulated.	<p>"I wish to read this book and miss the library."</p> <p>"That play had been recommended and I want to watch it."</p> <p>"Thanks the author to let us understand the history, a sad story, a grateful poem!"</p> <p>"There are a bunch of good anti-epidemic books waiting for me to read..."</p> <p>"Well done! Understanding the ancestors sacrificing themselves for our happiness today, we should be satisfied, grateful, learn their spirit for the revolution."</p> <p>"I learned that, thank you!"</p>
2	Readers were attracted by audio frequency, picture and text.	<p>"I come for the cover," "the audio reading is excellent, as if we are brought to the fish world."</p> <p>"Beautiful, beautiful pictures and texts."</p> <p>"The anchor's audio frequency is so great that It's like being there."</p> <p>"It is not easy to find an anchor who can recite the complicate book. The anchor is awesome."</p> <p>"I wonder what is BGM , and I read it at one sitting."</p> <p>"The anchor is very professional with solid broadcasting skills."</p> <p>"The anchor's voice has been echoing in my ears."</p>
3	Some readers left messages to express their love for the program.	<p>"RTLE leads us to start a cultural journey at home."</p> <p>"RTLE is my favorite program and I will continue reading all long."</p> <p>"Clear context, detailed content, this typesetting is really diligently."</p> <p>"The author writes so well."</p> <p>"The topic is very nostalgic to make me recall the story-telling scene when I was a child."</p>
4	There are more expectations.	<p>"RTLE as a whole should be more in-depth, not only in the aspect of book selection, but also the understanding of text and the reasons for recommendation."</p> <p>"I hope the name of the background music would be attached in the following program."</p> <p>"I always feel that the speed of the speech is a little bit fast."</p> <p>"The pronunciation is good, while the intonation lag behind."</p>

in terms of audio frequency, content, pictures, typesetting, etc., their wishes to read and acquire knowledge, their recommendations, and their expectations for a better column. The comment area of high-view tweets often becomes a place for communication and interaction between library and readers, and among readers themselves. For example, one issue of "Ugly or not is unimportant, whether I have personality matters!" which narrated Zhu Yuanzhang's magnificent entrepreneurial history, providing an image of grass-roots entrepreneurs, attracted the attention of many Ming Dynasty fans. Readers expressed their love and respect for the Ming Dynasty through their praise, left in messages on the platform. In addition, readers also recommended the classic book "Wanli Fifteen Years" of Ming history in the comment area, which provided reading guidance for other readers as well as a reference for library procurement. School alumni also left messages, "Readers graduated several years have even been reluctant to lose access to the library's public WeChat account." It can be seen from readers' positive feedback that the multisensory mode of reading promotion could attract reader attention and lead to additional reading. In the era of new media, readers are not only content readers, but also content creators. Readers' feedback is RTLE's greatest driving force for continuous growth and development, so librarians set up and published various comments for readers to discuss further to evoke their reading consciousness and promote deeper reading.

Analysis of multisensory reading

In the RTLE program, visual, audio, and emotional senses are motivated. How they work in the reading process have been analyzed in the following.

Analysis of visual reading

Rich visual input underlies reactivation of visual representation—that is, visual imagery—to generate reading comprehension. Importantly, the visual imagery evoked by the pictures and words is likely to set expectations in the reader's mind regarding the nature of the contents. For example, in the background of beautiful pictures of natural scenes, the 105th tweet introduced a picturesque place (Kekexili), successfully earning 523 page reviews. During the reading, visual imagery is also elicited. That is one might be filled with an awe for nature and want to consider the sensation transference that may occur, so that one expects what the contents of recommended book and the feeling of a real trip are.

Analysis of audio reading

Sound conveys the dynamics of the surrounding environment and provides information about its spatial, contextual, and high-level attributes to enhance the vividness of auditory imagery. The audio component of the program allowed readers to experience sound-induced cognitive processes and manipulate the

interaction between perceptual and contextual cues. On May 12th, the tweet of “Big love follows disaster, spring will eventually come” successfully earned 2048 page views, which described an event that Wenchuan earthquake in Sichuan province happened on the same day in 2008 and the huge impacts and rapid reconstruction of the area. As the master anchor’s sound and music build to a crashing crescendo, students can be easy to identify with the mood of the coming spring, after having suffered so much during the coronavirus pandemic and having been gradually recovering from it. Arguably, the program has spiritual value, allowing students to isolate themselves from disturbances and immerse themselves in the beauty of reading and relaxing their body and mind accompanying the broadcaster’s soft voice.

Analysis of emotional interactions

Multisensory reading allows us to mobilize the audiences’ perceptual senses, which is the simplest and most preliminary idea regarding it. Additionally, it needs to consider strategic research, communication and interaction in activities. In RTLE, readers share their inner feelings to respond to creators’ understanding and mood to induce emotional resonance. For instance, one alumnus left the message “*On campus, it is so quiet that I can hear my heartbeats and breathing sounds, feel birds flying and hares getting through the bamboo forest as well. It is the first time I feel that the mountain is more secluded while the birds are singing.*” Mental imagery triggered by multisensory inputs may play an important role in aiding the recall of the reader’s experience and account for the program’s emotional appeal.

Design and implementation of the RTLE program

From 2020 to 2021, the Zhejiang University of Technology Library launched its RTLE program on the library’s WeChat public account, divided into two parts: the “RTLE @ Librarians’ writing” column in 2020 and the “RTLE @ Readers’ Talk” counterpart in 2021. Aiming to recommend library collections to guide in-depth reading, the RTLE program constructs a promotional platform featuring multisensory promotion, multi-subject participation, and multichannel dissemination. Multilevel readers are recruited into RTLE working groups and divided into full- and part-time staff. Among these, librarians and students co-create the multisensory tweet contents and jointly run the program. After meticulous editing, while the tweets are released online, the quarterly audio comic books, compiled from the contents of tweets, are disseminated offline. The process of RTLE design is illustrated in [Figure 1](#) below.

Design—multisensory reading promotion

After accumulating a certain level of experience, library administrators decided to pour more inputs into this multisensory, interesting, interactive, and effective mode of reading promotion, going from producing one issue per week to producing one issue per day. RTLE advocates reading not just

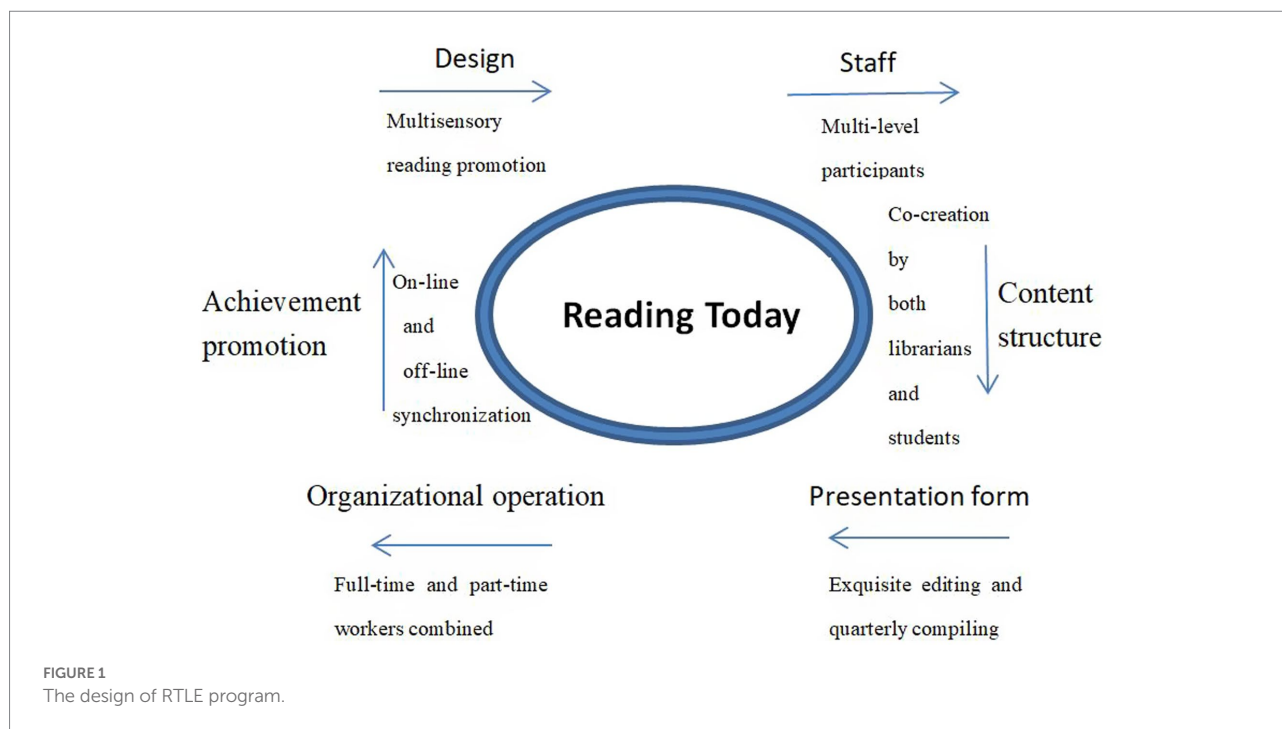
by seeing, but also by listening and through other multisensory ways. The program suggests an “audio plus pictorial” form of multisensory reading would provide readers with a better understanding of books. Relying on the social attributes of the WeChat platform, readers can naturally empathize with the spirit of works through their comments, messages, and discussions in comment areas to promote deeper reading. A complete closed loop of reading promotion would be created by librarians’ recommendations of books and readers’ comments on them. In a series of aural and pictorial tweets of RTLE, the authors recommend books through storytelling in an atmosphere rendered by music, human voices, pictures, and text elements.

Staff—multilevel participants

Characterized by flexibility and openness, RTLE is open to, and welcomes, anyone interested in the program at any time. In addition, program participants, according to their own advantages or wishes, can flexibly join one or more groups such as a “content-writing group,” an “anchor group,” an “editing group” and a “publishing group.” Program staff is composed of the library’s culture and broadcasting department, which is the full-time project component, managing program operation and personnel training; skilled and experienced librarians with excellent writing skills, recruited to ensure the quantity and quality of manuscript sources; winners of the prophase composition—writing or read-aloud contests enlisted on the team; teachers from the Humanities College invited to participate in program instruction and evaluation; broadcasting majors recommended by their professors to take on the post to gain experience; and students of the library association in charge of the WeChat editorial work. Among other things, ardent alumni and their children and friends have supported us by providing their read-aloud recordings and sharing their reading comprehension.

Content structure—co-creation by librarians and students

Cooperation between librarians and students enables the program to work steadily. The content structure includes text-based and audio-based portions, co-created by librarians and students, in the form of a combination of written and spoken words. In the text-based portion, most of texts were created by librarians, together with partial texts produced by outstanding students. Both librarians and students formed the “content-writing group,” the authors of which have various styles and wide interests to recommend library collections from different viewpoints. In the audio-based portion, students majoring or interested in broadcasting have been invited to make up the “anchor group” to complete sound recordings upon librarians’ training and text interpretation. Finally, the contents will be polished and verified several times by members of the “editing group” and the “publishing group” to ensure the release of high-value essays. RTLE has absorbed both the



reader group and the librarian group to form a major alliance of user-produced content represented by student readers and professional content producers represented by librarians.

Presentation form—meticulous editing and quarterly compiling

Following the completion of text and audio frequency, each tweet is rolled out after it is edited by work-group members, combining it with relevant pictures and information to increase reading interest and attention. Audio and video resources released on the library's public account have been stored on the WeChat platform for reader access at any time. Additionally, all materials are compiled in every quarter, including a quarterly disk of audio frequency and print booklet of text contents, which are preserved as references for subsequent activities. Although audio disks have not yet been collected in the library, they can be lent to support use when books are introduced inside and outside libraries. In addition, the library has also successively printed RTLE content into audio comic books as a supplementary material for offline reading guidance.

Organizational operation: A combination of full- and part-time workers

Program implementation and management are carried out jointly by librarians, students, teachers, alumni, etc. In terms of personnel arrangements, the RTLE program has set up the full-time three-librarian Culture and Broadcasting Department and recruited 165 part-time reading promoters. The three full-time librarians are responsible for project planning, preparation, implementation, evaluation, and

coordination and management of all part-time staff. The part-time team is composed of a variety of members, including 105 part-time student promoters (52 student anchors and 53 student volunteers), 48 librarians from other departments, and a total of 12 outside reading promoters (including two librarian friends, three children of alumni, one broadcaster from China Central Television, one chairman of the Disabled Persons' Federation, one retired teacher, and four rural commissioners). All part-time student promoters were grouped into shifts, with each group containing a leader (who could be in three to five groups), a typesetter, a copywriter and a photo editor.

Achievement promotion—on-line and off-line synchronization

The RTLE reading promotion program features integration of online micro-reading and offline deep reading. During the continuous push of the online WeChat column, relevant reading activities surrounding RTLE themes were regularly carried out, including extended online activities such as the classic read-aloud activity "Hearing Your Voice," the broadcasting drama "Long Word," the activity of new book recommendation "Getting to Know New Books," and supporting offline activities like reading conferences, book fairs, related lectures and essay contests, etc. In addition, the RTLE materials have been rearranged, compiled, and printed into a total of 1,200 audio comic books, divided quarterly (spring, summer, autumn and winter) to be donated and distributed to eight rural cooperative grid reading rooms to enrich the local reading room collection and help create a positive reading atmosphere.

Implications

Summarizing the relative activity experiences can help innovate mode of library reading promotion and explore much higher and deeper level of reader service. The RTLE program provides a multisensory reading experience by tapping into the potential of students' visual and audio abilities with pictorial and aural stories and essays. Moreover, the further interaction of commenters allows for emotional resonance and guides students to switch from shallow reading to deep reading. This enlightening and thought-provoking multisensory reading is not only considered as a new mode of reading promotion but also as an innovative academic library service. As mentioned in the previous part of the article, other libraries should give attention to the multisensory turn. For instance, [Bembibre and Strlič \(2017\)](#) analyzed samples from an old book which develop a 'historic book odour wheel'. And, as for many other odour/flavour wheels, the idea here is that this helps connect identifiable chemicals with people's reactions to them ([Bembibre and Strlič, 2017](#); [Spence, 2020](#)). Beyond the sound of the books themselves as they are interacted with, there is also the unique sound scope of the library. For example, in Duke Humfrey's library, a stream of the sounds of creaks, rustles, coughs, and traffic noises that you could normally hear could be released to help cope with the pandemic lockdown ([Kidd, 2020](#)). A growing recognition of the importance of the multisensory attributes of reading has led a number of libraries, museums, and archives to organize exhibitions that engage more visitor senses ([Classen, 2016](#)).

In the social background that sees the government actively promoting nationwide reading, the system-mature, widely point-distributed, and resource-rich public interest institution of the library has naturally become the core force of reading promotion. Its excellent cases and successful practices are certain to provide references for social reading ([Li et al., 2020](#)). In the present study, we found that multisensory reading attracted readers' attention, stimulated their reading interests, and promoted their reading practice. As the continuous development of social reading has spawned new reader groups and new reading service needs, multisensory participation in reading may provide a possibility to meet their needs with the assistance of modern technology. For instance, even if you do not sit down and read quietly, you can listen to the indirect or wonderful fragments of a book in a busy environment. Thus, integrating reading into a multisensory context can be extended to social reading to promote nationwide reading and construct a scholarly society, further enriching and expanding the theory of reading promotion and providing advice for administrative policy making. Similarly, a multisensory approach could also be of pedagogical benefits to arouse students' attention, increase learners' engagement in activities, and improve

students' academic outcomes. In pandemic times, the attention-arousing and interest-stimulating mode of multisensory reading in RTLE delivers a value of book with words and sound that can soothe and heal fear, anxiety, and other stressful emotions and solve the dilemma of college students' lack of spiritual guidance. Moreover, the present study hopes to help academic libraries to play their role in public emergencies.

It is important to mention some limitations of the present study. First, what we concluded from the present study based on the data of multisensory reading carried out in the RTLE program, in which readers are restricted to college students and the samples are small. Multisensory contexts are natural for life-long learning, so further research on big-sample and multi-level learners like schoolchildren, preschoolers, infants, and elderly people is necessary. Second, the multisensory presence of our study appears in the form of visual, audio, and emotional senses, which is somewhat different from a unisensory presentation. However, the addition of other perceptual modalities is supposed to flourish multisensory capabilities in future research, a relatively understudied topic. Third, the present study detected the effectiveness of multisensory stimuli in library reading promotion but did not observe differences between specific deviants like visual, audio, and emotional senses and evidence for implicit cross-modal correspondences. Nonetheless, it would be informative for future research to determine what might constitute an optional approach in terms of reading promotion in public or academic libraries and educational institutions.

Conclusion

The present study on reliable links between multisensory modes of reading promotion and students' reading tendencies cannot directly speak to their causality. Nonetheless, our results on the RTLE program would indeed suggest that multisensory inputs constitute an effective access to promote students' reading practice. They also support the applicability of multisensory modes of reading promotion to innovate library services. Although multisensory reading in the RTLE program remains a trial limited to visual, auditory, and emotional senses, its functions are certain to undergo further developments, along with the popularity of the program, the increasing growth of the WeChat platform, and the possibility of using other new media technology and the addition of other perceptual modalities. It would be particularly promising to apply multisensory reading promotion in other libraries, such as developing a new extension service and integrating multisensory engagement in social reading. Moreover, combined with a prompt administration time, the multisensory method can be an attractive potential didactic tool for student learning.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

YW: conception and design of the study, and writing—original draft. YJ: data curation and resource. WY: analysis of data and validation. YC: investigation and supervision. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by National Social Science Foundation of China (20BTQ028), Humanities and Social

Sciences Research Fund of the Chinese Ministry of Education (22YJC870010), and the Philosophy and Social Science Fund of Zhejiang Province (21NDJC039YB).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Aggleton, J. P., and Waskett, L. (1999). The ability of odors to serve as state-dependent cues for real-world memories: can Viking smells aid the recall of Viking experiences? *Br. J. Psychol.* 90, 1–7. doi: 10.1348/000712699161170
- Alexander, J. D., and Nygaard, L. C. (2008). Reading voices and hearing text: talker-specific auditory imagery in reading. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 446–459. doi: 10.1037/0096-1523.34.2.446
- Arora, S. (2021). Multisensory Experiences for Art Appreciation: Human-Computer Interaction – INTERACT.
- Bembibre, C., and Strlič, M. (2017). Smell of heritage: a framework for the identification, analysis and archival of historic odours. *Herit. Sci.* 5, 1–11. doi: 10.1186/s40494-016-0114-1
- Boerma, I. E., Mol, S. E., and Jolles, J. (2016). Reading pictures for story comprehension requires mental imagery skills. *Front. Psychol.* 7:1630. doi: 10.3389/fpsyg.2016.01630
- Bosman, R., Glover, J., and Prince, M. (2008). Growing adult readers: promoting leisure Reading in academic libraries. *Urban Libr. J.* 15.
- Bottini, G., Corcoran, R., Sterzi, R., Paulesu, E., Schenone, P., Scarpa, P., et al. (1994). The role of the right hemisphere in the interpretation of figurative aspects of language: a positron emission tomography activation study. *Brain* 117, 1241–1253. doi: 10.1093/brain/117.6.1241
- Brosch, R. (2018). What we 'see' when we read: visualization and vividness in reading fictional narratives. *Cortex* 105, 135–143. doi: 10.1016/j.cortex.2017.08.020
- Brunyé, T. T., Dittman, T., Mahoney, C. R., Walters, E. K., and Taylor, H. A. (2010). You heard it here first: readers mentally simulate described sounds. *Acta Psychol.* 135, 209–215. doi: 10.1016/j.actpsy.2010.06.008
- Chesnut, M. T. (2011). Reading between the lines: extending the reach of a university common Reading program via the academic library. *Pract. Acad. Librariansh.* 1, 51–58.
- Classen, C. (2016). *The Museum of the Senses: Experiencing Art and Collections*. New York: Bloomsbury Publishing.
- Dali, K., and McNiff, L. (2020). Reading work as a diversity practice: a differentiated approach to reading promotion in academic libraries in North America. *J. Librariansh. Inf. Sci.* 52, 1050–1062. doi: 10.1177/0961000620902247
- De Bruxelles, S. (2020). Scents and Sensibility! Books' Unique Smells Bottled. *Dly Mail*. Available from: <https://www.pressreader.com/uk/daily-mail/20200330/282226602823576> (Accessed March 30, 2020).
- Decety, J., and Ruby, P. (2001). Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nat. Neurosci.* 4, 546–550. doi: 10.1038/87510
- Doyle, M. C., and Walker, R. (2002). Multisensory interactions in saccade target selection: curved saccade trajectories. *Exp. Brain Res.* 142, 116–130. doi: 10.1007/s00221-001-0919-2
- Fajardo, A. (2010). Book clubs: not just for public libraries. *Coll. Undergrad. Lib.* 17, 65–69. doi: 10.1080/10691310903584783
- Gauder, H., Giglierano, J., and Schramm, C., Eds. (2007). Porch Reads: Encouraging Recreational Reading among College Students. Vol. 14. Colloque Louis Neel-couches Mince et Nanostructures Magnétiques, 1–24.
- González, J., Barros-Loscertales, A., Pulvermüller, F., Meseguer, V., Sanjuán, A., Belloc, V., et al. (2006). Reading cinnamon activates olfactory brain regions. *NeuroImage* 32, 906–912. doi: 10.1016/j.neuroimage.2006.03.037
- Gori, M., Ober, K. M., Tinelli, F., and Coubard, O. A. (2020). Temporal representation impairment in developmental dyslexia for unisensory and multisensory stimuli. *Dev. Sci.* 23:e12977. doi: 10.1111/desc.12977
- Hernandez, F. L., and de la Cruz, H. P. (2007). "365 days of books": a blog to promote reading. *Prof. Inf.* 16, 131–133. doi: 10.3145/epi.2007.mar.05
- Horbach, J., Scharke, W., Croll, J., Heim, S., and Günther, T. (2015). Kindergarteners' performance in a sound-symbol paradigm predicts early reading. *J. Exp. Child Psychol.* 139, 256–264. doi: 10.1016/j.jecp.2015.06.007
- Kidd, P. (2020). Coronavirus UK: In Tough Times, Brucie is Getting us Through. *Times*. Available at: <https://www.thetimes.co.uk/article/coronavirus-uk-in-tough-times-brucie-is-getting-us-through-2vr7bj8gd>. (Accessed April 11, 2020).
- King, A. J., and Lewkowicz, D. J. (2012). *The Developmental and Evolutionary Emergence of Multisensory Processing: From Single Cells to Behavior* Cambridge, MA: MIT Press.
- Kosslyn, S. M. (2005). Mental images and the brain. *Cogn. Neuropsychol.* 22, 333–347. doi: 10.1080/02643290442000130
- Kurby, C. A., Magliano, J. P., and Rapp, D. N. (2009). Those voices in your head: activation of auditory images during reading. *Cognition* 112, 457–461. doi: 10.1016/j.cognition.2009.05.007
- Lan, Y. Y. (2021). *Research on the reading promotion of Chinese traditional culture in double first class university libraries*. [master's thesis]. [Zhengzhou]: Zhengzhou University.
- Li, W., Yang, F., Mao, Y. Y., and Liu, Y. (2020). The roles of reading promoters in libraries: dimensions, antecedents and consequences. *J. Libr. Sci. China* 46, 73–87. doi: 10.13530/j.cnki.jlis.2020024
- Lim, M. (2015). Three-dimensional reading: a preliminary investigation on the mode of reading promotion activities in university libraries-taking the library of Hainan medical college as an example. *Sci. Ecol. Inner Mongolia* 6, 139–140, 142. doi: 10.3969/j.issn.1007-6921.2015.06.069
- Liu, M. (2019). Design of reading promotion model for college libraries based on immersion experience. *J. Libr. Sci.* 41, 64–68. doi: 10.3969/j.issn.1002-1884.2019.08.012
- Lv, H. H. (2021). Construction of the evaluation system of reading promotion with CIPP method: taking Shenzhen technology university as an example. *J. Acad. Libr. Inf. Sci.* 39, 36–41. doi: 10.3969/j.issn.1006-1525.2021.06.007

- Ma, Y. (2018). Research on the communication effect and the content analysis of WeChat reading promotion in: taking six fine art University libraries as examples. *Libr. Res.* 48, 1–8. doi: 10.3969/j.issn.2095-5197.2018.03.001
- Mackay, M., and Tarulli, L. (2015). Dal reads: the evolution of a university-based Unity Reading program. *Ref. User Serv. Q.* 54, 16–18. doi: 10.5860/rusq.54n3.16
- Matusz, P. J., Dikker, S., Huth, A. G., and Perrodin, C. (2019). Are we ready for real-world neuroscience? *J. Cogn. Neurosci.* 31, 327–338. doi: 10.1162/jocn_e_01276
- Moore, A. T., and Schwitzgebel, E. (2018). The experience of reading. *Conscious. Cogn.* 62, 57–68. doi: 10.1016/j.concog.2018.03.011
- Oxford's Bodleian Library. (2022) Available at: <https://www.bodleian.ox.ac.uk/bodley/using-this-library/rooms/dh>
- Paraskevopoulos, E., Chalas, N., Karagiorgis, A., Karagianni, M., Styliadis, C., Papadelis, G., et al. (2021). Aging effects on the neuroplastic attributes of multisensory cortical networks as triggered by a computerized music Reading training intervention. *Cereb. Cortex* 31, 123–137. doi: 10.1093/cercor/bhaa213
- Pereira, A. F., Smith, L. B., and Yu, C. (2014). A bottom-up view of toddler word learning. *Psychon. Bull. Rev.* 21, 178–185. doi: 10.3758/s13423-013-0466-4
- Perrone-Bertolotti, M., Kujala, J., Vidal, J. R., Hamame, C. M., Ossandon, T., Bertrand, O., et al. (2012). How silent is silent reading? Intracerebral evidence for top-down activation of temporal voice areas during reading. *J. Neurosci.* 32, 17554–17562. doi: 10.1523/JNEUROSCI.2982-12.2012
- Soliman, M. S., and Al-Madani, F. M. (2017). The effects of a multisensory based instruction combined with brain compatible environment techniques on Reading fluency and Reading comprehension of fourth-grade students with dyslexia. *Croat. J. Educ.* 19, 363–397. doi: 10.15516/cje.v19i2.2190
- Song, Y. (2022). *Research on the promotion strategy of experiential reading in university library*. [master's thesis]. [Changchun]: Changchun Normal University.
- Spence, C. (2020). The multisensory experience of handling and Reading books. *Multisens. Res.* 33, 902–928. doi: 10.1163/22134808-bja10015
- Spence, C., Pavani, F., and Driver, J. (2000). Crossmodal links between vision and touch in covert endogenous spatial attention. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 1298–1319. doi: 10.1037/0096-1523.26.4.1298
- The Eighteenth National Reading Survey Report. (2021). Available at: www.thepaper.cn/newsDetail_forward_12369739 (Accessed April 21, 2021).
- Warnick, K., and Caldarella, P. (2016). Using multisensory phonics to Foster Reading skills of adolescent delinquents. *Read. Writ. Q.* 32, 317–335. doi: 10.1080/10573569.2014.962199
- Weatherhead, D., and White, K. S. (2017). Read my lips: visual speech influences word processing in infants. *Cognition* 160, 103–109. doi: 10.1016/j.cognition.2017.01.002
- Wei, Q. Y., and Yang, Y. (2017). WeChat library: a new mode of mobile library service. *Electron. Libr.* 35, 198–208. doi: 10.1108/EL-12-2015-0248
- Yägüez, L., Nagel, D., Hoffman, H., Canavan, A. G. M., Wist, E., and Hömberg, V. (1998). A mental route to motor learning: improving trajectorial kinematics through imagery training. *Behav. Brain Res.* 90, 95–106. doi: 10.1016/S0166-4328(97)00087-9
- Yan, T. T. (2016). Investigation and analysis on the national wide Reading promotion project in Canada. *Libr. Dev.* 2, 77–80. doi: 10.3969/j.issn.1004-325x.2016.02.018
- Zhang, L. H., and Zang, H. M. (2018). Discussion on the innovative mode of experiential reading promotion in university libraries. *J. Libr. Sci.* 38, 47–48. doi: 10.3969/j.issn.1003-1588.2018.01.017
- Zheng, C. X. (2019). Research on the future transformation of university libraries under the phenomenon of continuous decline of paper books borrowing. *J. Libr. Sci.* 3, 75–78. doi: 10.3969/j.issn.1002-5197.2019.03.016



OPEN ACCESS

EDITED BY

Anil Ufuk Batmaz,
Concordia University,
Canada

REVIEWED BY

Marta Mendez,
University of Oviedo,
Spain
Giulia Cappagli,
Italian Institute of Technology (IIT), Italy

*CORRESPONDENCE

Xiaoming Tian
✉ tianxm@mail.usts.edu.cn
Bo Dong
✉ dongb283@126.com
Yuan Sun
✉ sunyuan0928@126.com
Xiuling Zhang
✉ journeyonstars@126.com

[†]These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION

This article was submitted to Perception Science, a section of the journal Frontiers in Psychology

RECEIVED 05 October 2022

ACCEPTED 16 December 2022

PUBLISHED 11 January 2023

CITATION

Dong B, Chen A, Gu Z, Sun Y, Zhang X and Tian X (2023) Methods for measuring egocentric distance perception in visual modality.
Front. Psychol. 13:1061917.
doi: 10.3389/fpsyg.2022.1061917

COPYRIGHT

© 2023 Dong, Chen, Gu, Sun, Zhang and Tian. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Methods for measuring egocentric distance perception in visual modality

Bo Dong^{1*†}, Airui Chen^{1†}, Zhengyin Gu², Yuan Sun^{3*},
Xiuling Zhang^{4*} and Xiaoming Tian^{1*}

¹Department of Psychology, Suzhou University of Science and Technology, Suzhou, China,

²Department of Psychology, Zhejiang Sci-Tech University, Hangzhou, China, ³School of Education, Suzhou University of Science and Technology, Suzhou, China, ⁴School of Psychology, Northeast Normal University, Changchun, China

Egocentric distance perception has been widely concerned by researchers in the field of spatial perception due to its significance in daily life. The frame of perception involves the perceived distance from an observer to an object. Over the years, researchers have been searching for an optimal way to measure the perceived distance and their contribution constitutes a critical aspect of the field. This paper summarizes the methodological findings and divides the measurement methods for egocentric distance perception into three categories according to the behavior types. The first is Perceptual Method, including successive equal-appearing intervals of distance judgment measurement, verbal report, and perceptual distance matching task. The second is Directed **Action** Method, including blind walking, blind-walking gesturing, blindfolded throwing, and blind rope pulling. The last one is Indirect **Action** Method, including triangulation-by-pointing and triangulation-by-walking. In the meantime, we summarize each method's procedure, core logic, scope of application, advantages, and disadvantages. In the end, we discuss the future concerns of egocentric distance perception.

KEYWORDS

egocentric distance perception, spatial perception, measurement methods, perceptual method, direct action method, indirect action method

1. Introduction

Space is the basis of the interaction between human beings and the environment. “How do human beings perceive space?” is a classical problem in cognitive psychology and ecological psychology. With the emergence of interdisciplinary studies, the problem constitutes the foundation of environmental psychology. The “mystery” of spatial perception is that the visual system can reproduce three-dimensional space depending on two-dimensional retinal images (Gibson, 1950). The transformation mentioned above can be called depth perception. Specifically, depth perception includes exocentric distance perception, which is about the distance from one object to another, and egocentric distance perception, which is about the distance from an observer to an object (Blake and Sekuler, 2006; Ooi and He, 2007). Egocentric distance perception supports humans in representing

the location of objects and constructing the visual space of the environment (Gibson, 1950; Loomis et al., 1996). When humans take action in daily life, for example, taking a cup, driving a car, or throwing a draft, egocentric distance perception plays an essential role in the procedure. To some extent, egocentric distance perception is a precondition for survival.

According to the relative motion state between the observer and the object, egocentric distance perception can be divided into static state and dynamic state (Blake and Sekuler, 2006). For the static state, the observer and the target keep still, and for the dynamic state, the target or the observer is in motion. Egocentric distance perception in the dynamic state involves complex cognitive procedures, for example, motion parallax, high-speed spatial updating, speed perception, and ontological motion. Due to the complexity, there were few studies about it, and the study of its internal mechanism was still in infancy (Santillán and Barraza, 2019; Dong et al., 2021a). By contrast, researchers have studied egocentric distance perception in the static state more profoundly, and more than 10 methods have been developed, which have been widely used in laboratory research and business development. Furthermore, in essence, static distance perception is the basis and premise of dynamic distance perception, so the following will focus on egocentric distance perception in the static state.

Regarding egocentric distance perception, the problem that first needs to be solved is how to measure it; in other words, it is about the methods for egocentric distance perception (Loomis and Philbeck, 2008). The unique characteristics of egocentric distance perception determine the importance of measurement methods. Individuals can perceive the distance accurately, but they are disabled to speak the distance through introspection. Such as, people can easily and accurately take up a cup, but it is hard to say how far away it is. Some argued that egocentric distance perception was an automatic or unconscious perceptual procedure (Blake and Sekuler, 2006). Because of this, verbal report only partially reflects egocentric distance perception and lacks precision (Philbeck and Loomis, 1997). An optimal method for egocentric distance perception is action, such as walking to the target or touching the target (Philbeck and Loomis, 1997).

Nevertheless, researchers also need different methods to choose from because of the diversity of distance ranges. Cutting (1995) divided the space around individuals into the private space (0–2 m), the action space (2–30 m), and the vista space (>30 m). The visual system utilizes the binocular parallax to estimate egocentric distance in the private space. Humans can touch objects directly with accuracy. When egocentric distance is beyond 2 m, namely in the action space or in the vista space, it is difficult to show the perceived distance by hand (Bufacchi and Iannetti, 2018). This suggests that researchers need to find other ways to measure egocentric distance perception accurately.

In sum, a systematic understanding of the similarities and differences between the measurement methods is an important guarantee for further exploration of egocentric distance

perception. In the action space (2–30 m) and the vista space (>30 m), the researchers measured the estimated distance in the static state by a large number of methods, such as successive equal-appearing intervals distance judgment measurement (Gilinsky, 1951; Ooi and He, 2007), verbal report (Philbeck and Loomis, 1997), perceptual distance matching (Wu et al., 2007a), blind walking (Thomson, 1983), blind-walking gesture (Ooi et al., 2001), blindfolded throwing (Eby and Loomis, 1987), blind rope pulling (Philbeck et al., 2010; Bian and Andersen, 2013), triangulation-by-pointing (Fukushima et al., 1997), and triangulation-by-walking (He et al., 2004). Studies showed that these methods could measure relatively pure perceived distance. The relationship between the measured distance and the practical distance fits the linear function, and there is also a significant positive correlation between the measurement results of different methods (Loomis et al., 1992). However, there are also significant differences in the measurement results between different methods.¹ The differences also exist in the core logic of measurement, the scope of application, the advantages, the disadvantages, and the precision of the methods (Hutchison and Loomis, 2006; Bian and Andersen, 2013). Thus, our paper according to the type of action divides the methods for egocentric distance into Perceptual Method (not relying on action), Direct Action Method (relying on action), and Indirect Action Method (relying on indirect action). At the same time, the core logic, the measurement procedure, the scope of application, the advantages and the disadvantages of the methods are summarized, to provide a reference for the researchers in the field of spatial perception.

2. Perceptual method

Early researchers believed that the egocentric distance could be realized or spoken by verbal reporting or comparing (Loomis et al., 1992). This is a simple and direct way to measure. In this paper, we name the kind of methods 'Perceptual Method'. It is an early kind of method used in the research of egocentric distance perception, which has the following three characteristics. Firstly, in the process of estimating the distance, it is no need for observers to move. Secondly, judged or reported distance by observers means their perception. Finally, when observers are reporting, there is still continuous visual input. This kind of method involves Successive equal-appearing intervals of distance judgment measurement, Verbal report, and Perceptual distance matching.

¹ There is no contradiction between positive correlation and significant difference between different paradigms. Such as, there is a positive correlation between two groups of data ([1,2,3] and [8,10,12], $r=1.00$, $p<0.001$), but there are still significant differences between them ($t(2)=13.86$, $p=0.0052<0.01$).

2.1. Successive equal-appearing intervals of distance judgment measurement

An important task of psychology is to describe the relationship between physical quantity and psychological quantity. Gilinsky (1951) developed successive equal-appearing intervals of distance judgment measurement to establish a functional relationship between the physical quantity and the psychological quantity of egocentric distance (Gilinsky, 1951; Ooi and He, 2007). In the experiment of Gilinsky (1951), the observer needed to instruct an experimenter to adjust the distance between a horizontal rod stick and a pointer stick on the ground until the distance equated to 1 foot in the memory of the observer. Throughout the experiment, the observer was required to instruct the experimenter to adjust the intervals from near to far on the ground multiple times (Figure 1).

The following three points are the core logic of the method: (1) Distance in the observer's memory is a psychological quantity; (2) The memory of distance can be extracted; (3) The psychological quantity of egocentric distance can be added but not multiplied, namely in terms of the psychological quantity, 2 feet is the sum of the 0 to 1 foot and the 1 to 2 feet, not twice the 0 to 1 foot. In practical use, if the psychological quantity of a standard distance (e.g., 1 foot) can be adjusted continuously at equal intervals on the ground, the psychological quantity of any distance can be calculated. Using the method, Gilinsky (1951) found that the farther the location was, the longer 1 foot in the observer's

memory was. Egocentric distance (the physical quantity) and estimated distance can fit a specific equation, namely Gilinsky Equation.

$$d = \frac{DA}{D + A}$$

d is the psychological quantity of egocentric distance; D is the physical quantity of egocentric distance; A is a constant, representing the scale value of the distance. The value of A depends on the observer and depth cues in the current environment.

Successive Equal-appearing Intervals Distance Judgment is one of the earliest experimental methods to study egocentric distance perception by psychophysical method (Gilinsky, 1951; Ooi and He, 2007). The results of the method not only showed that there was a correspondence between physical quantity and psychological quantity but also established the function between the two (Gilinsky Equation). However, there are two deficiencies in the method. In the first place, the psychological quantity of the first 1 foot relies on the observer's memory, and the individual difference will affect the result. Given the problem, Ooi and He (2007) revised the method. In their experiment, researchers first placed a graduated tape measure in front of the observer. Then they would ask the observer to look at the graduated position of a distance, such as 2 feet, and to memorize the distance from the tote to the position as a reference for the subsequent

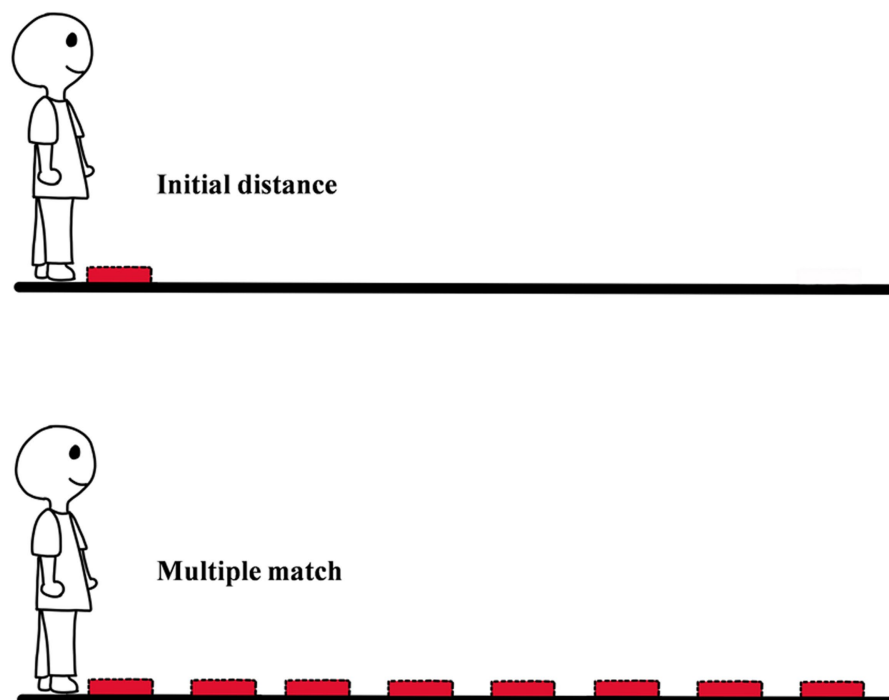


FIGURE 1
The diagram of successive equal-appearing intervals distance judgment measurement.

distance (Ooi and He, 2007). In the second place, starting from the second interval, the distance in the method is the exocentric distance instead of the egocentric distance. This leads to the calculated psychological quantity of 1 foot involving both egocentric distance and exocentric distance; for example, estimated 3 feet = egocentric distance (0 to 1 foot) + exocentric distance (1 to 2 feet) + exocentric distance (2 to 3 feet). In addition, the longer the distance, the greater the effect of exocentric distance. This flaw may be one of the reasons why the distance compression in the Gilinsky Equation is much larger than in other studies (Loomis et al., 1996). For example, the studies that used blind walking found that observers could accurately perceive egocentric distance out to 25 meters. However, when using successive equal-appearing intervals distance judgment, the observer could only perceive the distance accurately out to 5 meters. Because of this, in recent studies, researchers little used the method. Nevertheless, the Gilinsky equation revealed an important and still closely watched phenomenon in the field of distance perception and spatial perception-perceptual compression of distance. Therefore, researchers still often mentioned the method in recent papers.

2.2. Verbal report

Verbal report is that when an observer estimates egocentric distance, he/she gives a quantitative estimate of the distance in specific units, and in some studies, it was called “Magnitude Estimation” (Loomis et al., 1992; Philbeck and Loomis, 1997; Loomis and Philbeck, 2008). Specifically, the procedure of verbal report is relatively simple. The first is that observer looks at the target on the ground and tries to estimate the distance between himself/herself and the target. Secondly, the observer needs to report the result of estimating in a specific unit (e.g., feet, inches, yards, meters, centimeters, etc.) and the estimation can be regarded as the psychological quantity of egocentric distance (see Figure 2; Wu et al., 2007b). Verbal report is also an early method that was used to measure egocentric distance perception. The core logic of the method is that the perception of the observer is at the level of consciousness, and it can be spoken out. The distance reported by the observers means their naive cognition of distance. Because in daily life, people often use oral ways to express their cognitive results.

Compared to the successive equal-appearing intervals of distance judgment measurement, verbal report is a relatively simple and direct method that is not disturbed by exocentric distance perception (Wu et al., 2007b). In addition, verbal report can be used to measure distance perception under different conditions in a short time. Due to the advantages above, the method has been widely used in the domain of ground representation, spatial training, traffic safety, and others (Cavallo et al., 2001; Wu et al., 2007b; Bernhard et al., 2021). However, the method has also been challenged on two points over the years. The first is subjective cognitive correction. Adults are generally

aware of how their perception varies with distance, such as the principle “the farther the distance is, the smaller the object is.” Therefore, when adults report the estimation of distance, they may intentionally or unintentionally correct the result. To solve the problem, some researchers have tried to make observers report apparent distance instead of objective distance that was consciously corrected (Loomis et al., 1992). However, in the method observer has to measure the apparent distance in a unit. It is difficult to eliminate cognitive correction from the measurement procedure. Furthermore, the apparent distance is variable, which can also make the measurements unstable. Second, the core logic of the method is that perception can be spoken out. The logic is not entirely correct. As mentioned above, it is hard to speak out exactly how far away the cup is in daily life. Even though, before reporting people carefully observe the distance, they would also question their answers. It is worth noting that the sense of doubt and uncertainty does not mean that people cannot accurately perceive the distance. For example, a person can easily and accurately pick up a cup even with his/her eyes closed. It means that the perception of people is accurate. Although, like other measurement methods, the perceived distance measured by verbal report is smaller than the physical quantity, the distance reported by observers is significantly smaller than that measured by other methods (e.g., blind walking; Napieralski et al., 2011). At the same, verbal report is less reliable and less stable compared to other methods (Rand et al., 2019). This means that the reported distance is only part of the perceived distance, not all of it.

Because of its convenience and understandability, verbal report is still one of the commonly used methods in physical distance research. Different from the early studies, when researchers recently used verbal report to measure, they would supply other methods to cover the shortage of verbal report.

2.3. Perceptual distance matching

Perceptual distance matching is the method of showing egocentric distance perception of the observer by moving a matching target (Sinai et al., 1998; Wu et al., 2007a). To be more specific, in an experiment, after the observer views the target and remembers egocentric distance, he/she can turn around by 90° or 180° and instruct the researcher (or use remote control) to adjust the matching target on the ground until it is at the same distance as the target. In the process of adjusting, the observer can turn around multiple times to confirm whether the distance of the test target and the distance of the matching target are equal and adjust the matching target until satisfied without a limit of time (see Figure 3). Finally, the distance between the matching target and the observer can be regarded as the perceived egocentric distance of the target.

The core logic of Perceptual distance matching is setting the matching target. Although the method still assumes that perceived distances are in the individual's consciousness, the setting of the

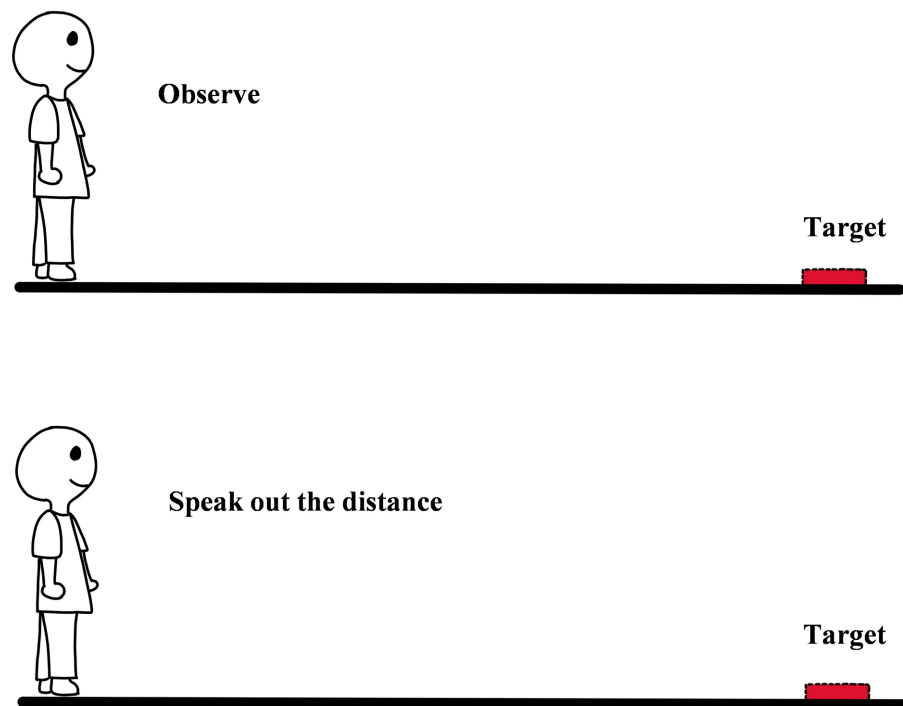


FIGURE 2
The diagram of verbal reporting.

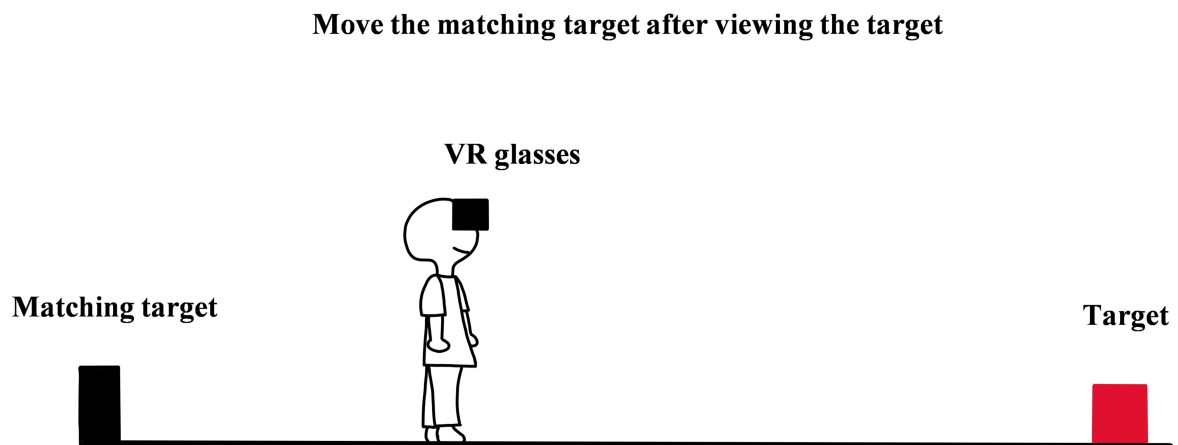


FIGURE 3
The diagram of perceptual distance matching (in virtual reality).

matching target transforms the perception of an absolute distance into a comparison of two absolute distances. In the comparison, the observer does not have to report the distance in a specific unit and just needs to adjust the location of the matching target to make the two distances look equal. Because of this, the method avoids the bias caused by prior knowledge of distance. To a large extent, the two deficiencies of verbal report, which are cognitive correction and “perception can be spoken out,” can be overcome.

Due to the principle of near is bigger and far is smaller, namely size-distance invariance, the estimation of distance would transform into the estimation of size when the matching target and the target are in the same field of vision. To avoid the effect of size, the experimenter generally would set the matching target at a different angle (usually on the opposite side of the target), so that the observer must turn around to see the matching target. A study showed that there was no significant difference between the

matching distance (3.54 ± 0.07 m) and the physical distance (3.66 m; [Sinai et al., 1998](#)). This means that the method is effective and accurate.

Perceptual distance matching does not depend on the verbal report or the direct action and has a wide scope of applications. On the one hand, because of the matching target and the target on the different sides, the method is easy to use in the field which is inconvenient to walk, for instance, on the mountain, water, and ground with pits or convex. On the other hand, during matching, the observer does not need to use the absolute or actual units of length, so the method can be used in a context with large systematic errors, and to some extent, the errors can be offset in the process of matching. For example, in virtual reality, the bias of distance perception is large, but perceptual distance matching is still appropriate for the scenario ([Wu et al., 2007a](#); [Dong et al., 2021a, b](#)). More importantly, neither the verbal report nor the action-based method cannot be applied in the vista space (>30 m), but perceptual distance matching has no such limitation.

2.4. The summary of perceptual method

In short, all the three methods above are the traditional methods in the studies of egocentric distance perception, and the observer expresses the psychological quantity of distance through verbal reporting or distance matching. The advantages of the perceptual method are simple principles, direct measurement, and easy to understand. The disadvantage is that it is difficult to exclude the influence of subjective cognitive factors. When using the method, the observer is always with visual input. Because of this, it is easy to correct the perceptual result. This makes the inherent mechanisms of the three perceptual methods complex, and slight factor variation may lead to a change in the measured distance perception.

3. Direct action method

As the understanding of perceptual paradigms deepened, researchers found that the action of body moving could more accurately and directly show egocentric distance, so they developed several measurement methods that are still widely used today. In this paper, the kind of paradigm is called the direct action paradigm, including blind walking, blind walking gesturing, blindfolded throwing, and blind rope pulling. The characteristics of the direct action paradigm are as follows: (1) The observer expresses the psychological quantity of egocentric distance by action rather than verbal reporting, (2) The observer expresses the distance to the target by approaching it in a straight line, which does not involve the change of angle ([Thomson, 1983](#); [Eby and Loomis, 1987](#); [Loomis et al., 1992](#); [Fukushima et al., 1997](#); [He et al., 2004](#); [Bian and Andersen, 2013](#)).

3.1. Blind walking

Blind walking is widely used in the studies of egocentric distance perception, in which the observer carefully observes and remembers the location of the target and then walks to the location of the target with eyes closed or a blindfold (see [Figure 4](#); [Thomson, 1983](#); [Philbeck et al., 2010](#); [Teng et al., 2016](#); [Zhou et al., 2016](#); [Li, 2017](#); [Siegel and Kelly, 2017](#); [Kelly et al., 2018](#)). The method was developed by [Thomson \(1983\)](#). His experiment was carried out in a flat space. The experimenter set up two pathways 6 meters apart on the ground, and the target was placed about 1 meter to the left of the walking path. The task of the observer was to walk toward the target with his/her eyes closed and place his/her foot as close as possible to the center of the target. To avoid flaws of the method, [Thomson \(1983\)](#) did his best to deprive the vision. In his preliminary experiment, he tried a variety of eye masks, but he found the observers could not walk naturally at a reasonable speed. So instead of eye masks, he asked participants to close their eyes to deprive the visual input. Before the formal experiment, participants were generally instructed to practice several times to walk naturally with their eyes closed. In the formal experiment stage, the experimenter would put the target in the specified position and ask the observer to watch the target for 5 s. And then, the observer closed his eyes and walked toward the target. After the observer stopped walking, he/she still had to close eyes and wait for the experimenter to mark where the observer stopped and move the target to the other position for the next trial. At the same time, another experimenter would take the observer back to the starting point still with eyes closed to avoid the observer receiving information about how well he/she had done in the previous trials (i.e., no feedback; [Thomson, 1983](#)).

The core logic of blind walking is reflecting perceptual performance by walking. The method involves not only the perceptual process of distance but also the memory of the initial location of the target, proprioception, and spatial updating ([Loomis et al., 1996](#)). The method has three advantages: (1) The method does not depend on the consciousness of the observer and is not affected by the subjective cognitive correction. When using the method, it is no need for observers to use any units of length (i.e., meters, feet, etc.) as a reference. They just need to remember the apparent distance. So blind walking can be performed even if the observers do not recognize or remember any units of length. (2) It is more reasonable for observers to show perceived distance by walking. The default unit for walking is the length of the step. As human beings need to walk every day and always measure the distance of objects around them by step, step becomes the most familiar unit of length for humans. (3) Walking with eyes closed avoids the feedback. Compared to the continuous vision input of Perceptual Method, blind walking shields the visual stimulus during walking. Therefore, the method further avoids observers optimizing the perceptual results while walking.

In addition to traditional blind walking, the researchers have developed some variations while keeping the core logic: (1) People can walk blindly on a treadmill instead of on the ground. This task



is called blind treadmill walking (Proffitt, 2006; Bossard et al., 2020). (2) Walking can also be done instead of imagining. The experimenter first measured the observer's conventional walking speed and then asked the observer to imagine walking to the target with his eyes closed. After recording the walking time, egocentric distance perception could be obtained by multiplying the walking speed. The task is called timed imagined walking (Plumert et al., 2005).

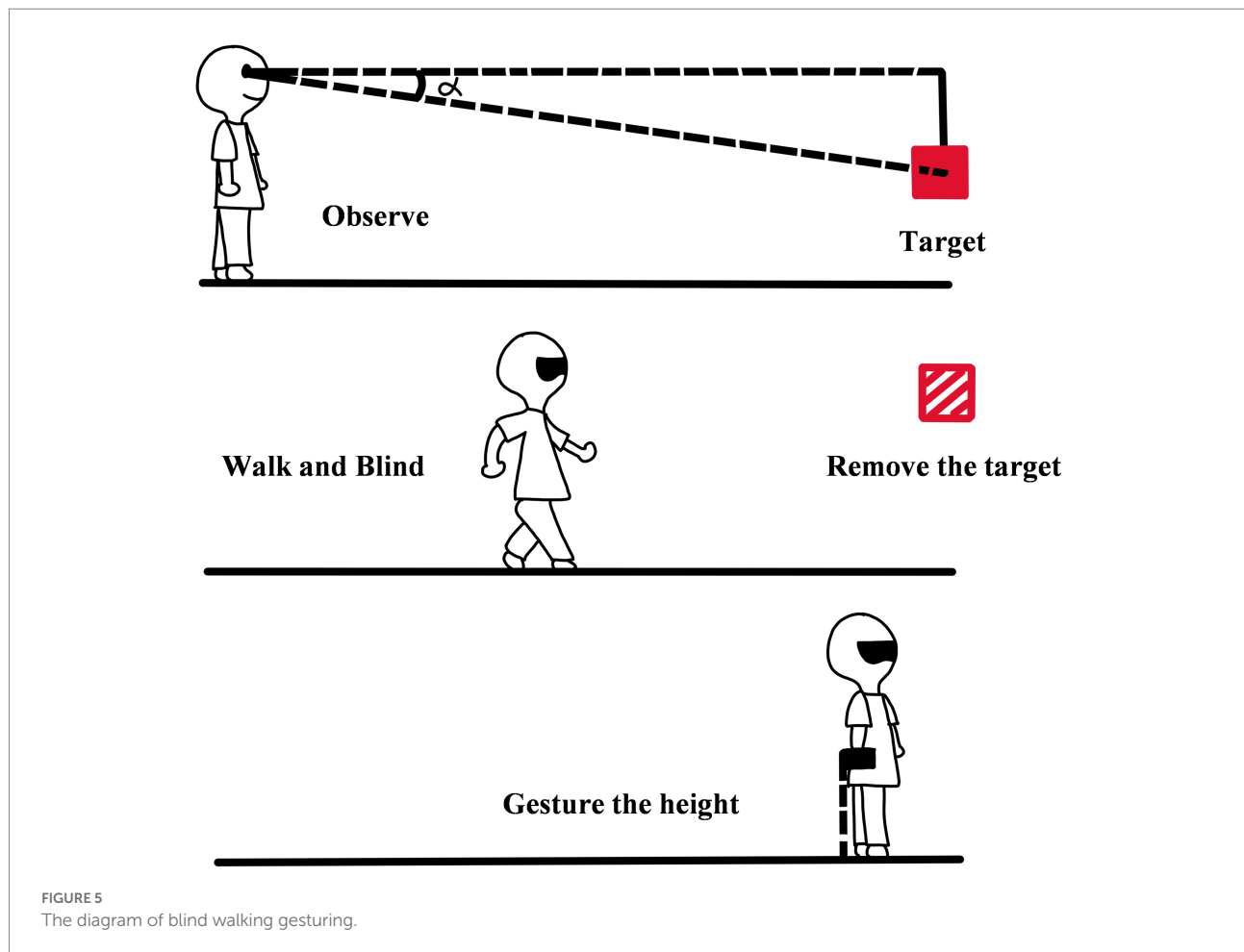
Blind walking is also restricted by several factors. The first is walking time. Thomson (1983) found that the duration of the test to finish the task would affect their performance. If observers finished the task within 8 s, they were able to accurately perceive egocentric distance within a range of 21 meters. If not, the accuracy of perception would decrease rapidly. This may be due to the limited holding time of the motion program (Thomson, 1983). The second is the need for space. Blind walking has high requirements on the space of the experiment field. There is plenty of space outside, but it is easily disturbed by bad weather; Indoor can eliminate the influence of the weather, but it is easy to collide with the objects around because of the narrow corridor. Some tried to use blind walking in virtual reality, but the limitation of devices could result in insufficient measuring range or a large number of system errors (Philbeck et al., 2010; Li, 2017; Siegel and Kelly, 2017; Kelly et al., 2018). The Final is the task load and the time limit of the whole experiment. Blind walking takes time and effort. It is difficult for the aged and the brain injury patient to

finish the task (Bian and Andersen, 2013; Wallin et al., 2017), and in extreme cases, participants can get hurt, so blind walking may not be the best choice when studying special populations (elders or patients).

3.2. Blind-walking gesturing

Blind-walking gesturing refers to the method in which observers first view the target and judge distance and height, then walk blinding according to the distance of the target in memory, and show the perceived height of the target with gesturing after reaching the location of the target (see Figure 5; Ooi et al., 2001; Chen et al., 2016). Blind-walking gesturing is a classical variant of blind walking, and the gesturing height is added to represent the perceived vertical height of the target (Ooi et al., 2006). The specific procedure of the method is as follows.

There are two stages in the method, including the perceiving stage and the reacting stage. In the perceiving stage, the observer is asked to observe the target and remember its egocentric distance. After finishing the observation, the observer should wear a blindfold, and this means that the perceiving stage is over. And then entering the reacting stage. At the time, the experimenter will quickly take the target away and instruct the observer to walk to the remembered location of the target. After the observer arrives, he/she will gesture the height of the target with hand. There is no



feedback in each trial. The method can not only measure the accuracy of distance judgment but also calculate the perceived direction of the target by gesturing height. The indicator of the perceived direction in the method is the angular declination below the horizon, which is represented by α . The calculation formula of α is $\alpha = \arctan [(h_{eye} - h) / d] \times 180 / \pi$ (h_{eye} is the eye height of the observer, h is the actual height of the target, d is the actual distance from the target to the observer). The psychological quantity of α is α' , and the calculation formula of it is $\alpha' = \arctan [(h_{eye} - h_g) / d_w] \times 180 / \pi$ (h_g is the height of the observer gesturing, d_w is the horizontal distance of the observer walking blind). By comparing α with α' , the accuracy of the observer's direction judgment can be obtained. Different from blind walking, the distance from the eye to the target is used as the indicator of distance estimation in the method. The calculation formula of the indicator is $d_{eye-to-target} = \sqrt{[d_w^2 + (h_{eye} - h_g)^2]}$ ($d_{eye-to-target}$ is the distance from the observer's eyes to the target, the meanings of other letters are consistent with the formula of α and α' (Ooi et al., 2001; Chen et al., 2016).

Blind-walking gesturing is added to the measurement of the target height based on blind walking; therefore, the research scope is no longer limited to the distance judgment on the ground and can be extended to the direction and height of the suspended

target. Some found that in the dark, the observer would underestimate egocentric distance, but they could accurately perceive the direction of the target (Thomson, 1983; Ooi et al., 2001, 2006).

The innovation of the method is that it extends the perceptual representation limited in the two-dimensional plane to the three-dimensional space and integrates the distance and the direction judgment into a method. It enables researchers to obtain more data in a study, which not only improves the research efficiency but also enables researchers to study spatial perception more comprehensively.

3.3. Blindfolded throwing

Blindfolded throwing refers to the method in which the observer throws an object (such as a bean bag) with eyes closed or covered after observing the location of a target (see Figure 6; Eby and Loomis, 1987; He et al., 2004; Peer and Ponto, 2016). The method is derived from the study of Eby and Loomis (1987). In the method, the observer first constructs the representation of egocentric distance to the target. And then, he/she programs an action based on the representation. Finally, the observer acts

(e.g., throwing a bean bag) with his eyes closed to show the perceived egocentric distance (Eby and Loomis, 1987). The method is similar to blind walking. The throwing distance can be regarded as the egocentric distance perception of the observer, and the performance of throwing is equal to the accuracy of egocentric distance perception. In the field of distance perception, the method is not commonly used, but sometimes, it would be used to explore interesting theoretical questions. For example, He et al. (2004) used the method to verify the sequential-surface-integration- process (SSIP) hypothesis (He et al., 2004). In the frame of SSIP, an obstacle on the ground between the observer and the target would make the observer underestimate the egocentric distance to the target. To verify the hypothesis, He et al. (2004) designed an experiment with two conditions. One condition was with an obstacle on the ground, and another was without an obstacle on the ground. The tasks in the two conditions were the same. The observer needed to view the target and judge egocentric distance. And then he/she threw a bean bag to the location of the target. The result showed that compared with the non-obstacle condition, the throwing distance of the observers under the obstacle condition was shorter. This means that SSIP is true (He et al., 2004).

The limitation of the method is obvious. The accuracy of it is lower than that of blind walking, and the performance of throwing is easy to be affected by the ability of the individual and practice. Eby and Loomis (1987) found that their participants were able to accurately throw with eyes closed, even when the distance was 15 m, but in the experiment of He et al. (2004), their participants could not do the same things. By comparison, the participants of Eby and Loomis (1987) had stronger athletic ability and stronger throwing ability than those of He et al. (2004). This means that the

motion of the observer needs to be balanced in advance when using the blindfolded throwing. In addition, He et al. (2004) also found that the feedback in the practice stage was beneficial to the performance of the participants. The foundation is in line with Eby and Loomis (1987), so whether the feedback is given is also an important factor that should be under consideration in the method.

3.4. Blind rope pulling

Blind rope pulling refers to the method in which after observing the target, an individual imagines that the target is tied to a rope and pulls the rope with eyes closed until the imagined target is pulled to the hand (see Figure 7; Philbeck et al., 2010; Bian and Andersen, 2013). In some cases, the rope can be used instead of the tape measure to measure easier. Philbeck et al. (2010) developed blind rope pulling based on blind walking. The procedure of measurement is as follows.

During measurement, the observer first practiced pulling the rope and formed knowledge of the speed and the length of the rope. And then, he/she observed the target and its location until being ready. After observation, he/she needed to close the eyes and prepare to pull the rope. During pulling, an experimenter would stand at the location of the target and place a tape (easier to measure the distance) around his waist. The experimenter should remember the tape measure scale (point A) where the hands are, and the zero which is the end of the tape measure was held by the observer. After the experimenter gave the signal, the participant started to pull the tape measure until the distance pulled was equal to the perceived egocentric distance. During the stage, the

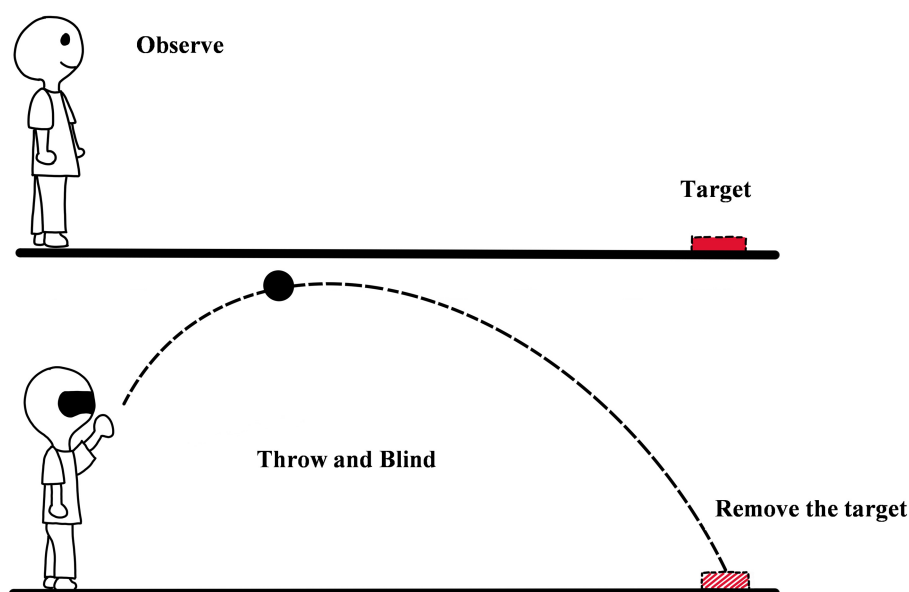
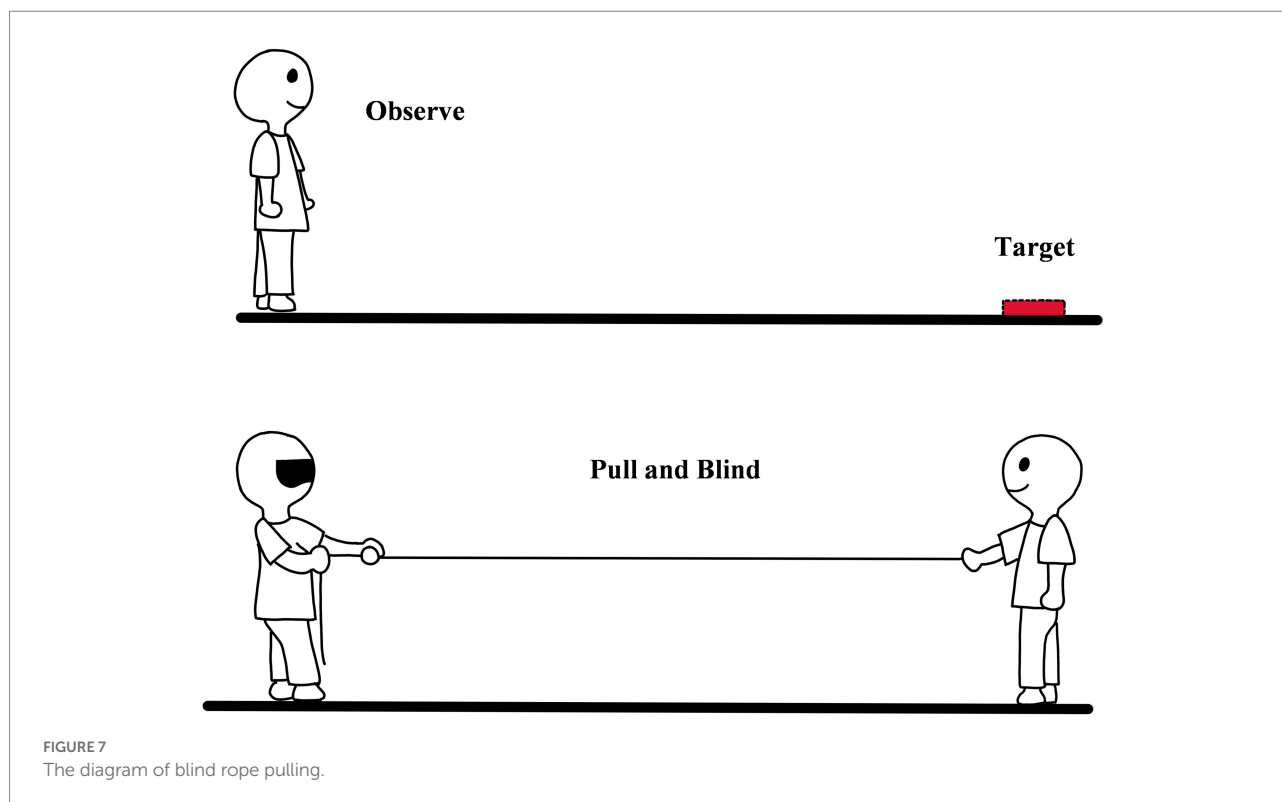


FIGURE 6
The diagram of blindfolded throwing.



experimenter let the tape measure pass freely. When the participant reported that the pulling was over, the experimenter would record the scale (point B) of the tape measure in hand. The perceived egocentric distance could be measured by B-A. To get more data, the time of pulling and the number of pulling times could also be recorded (Philbeck et al., 2010). There is a high correlation between the results of blind rope pulling, real egocentric distance, the results of perceptual method, and blind walking, which indicates that the method could measure egocentric distance perception (Philbeck et al., 2010; Bian and Andersen, 2013).

Eye closing and pulling rope are the two core settings of the method. Before the development of the method, the researchers had used another two methods. One is 'pass the rope through the subject's hands without arm movement' (Mershon et al., 1977) and the other is 'pull the rope with eyes opened' (Hagen and Teghtsoonian, 1981), but these methods lack precision. The core logic of blind rope pulling is similar to blind walking. The former uses the familiar arm length as a reference, while the latter uses the familiar step length as a reference, both of which reflect humans using the body to perceive the outside world. Before the formal experiment, observers would be asked to perform rope-pulling exercises with visual stimulation, which could effectively improve the accuracy of the task. And observers can calibrate their movements by visual rope, consciously adjust their movements, and reduce the errors caused by improper action. Compared to blind walking, the method is also easy to be affected by the design of the experiment. In a complex experimental design, its performance was different from that of blind walking (Philbeck

et al., 2010). There are two distinct advantages of blind rope pulling over the blind walk: (1) The method has lower requirements on the environment, and the scope of application is wider. It can be used not only on flat land, but also on the sea, lakes, and even on the edge of cliffs. Due to the small requirement of space for pulling and the experimenter who holds the rope being not necessary, even in a narrow space, the method still can be applied. It is worth noting that the method can be used to study not only the distance perception all around but also the distance perception above and below the observers. (2) Since the method does not require the ability of walking, it is more suitable for studying the distance perception of the elder and the brain injury patient.

3.5. The summary of direct action method

The direct action method was most frequently used in recurrent studies. The similarity of the direct action method is that the subjects show the perceived egocentric distance directly through action, relying on the perceptual representation or maintained memory after visual deprivation (Bian and Andersen, 2013). Some believed that these tasks reflect the 'post perception rather than distance perception' (Proffitt, 2006; Hajnal et al., 2016). However, other researchers have empirically proved that even when the stimulus has disappeared, the observer still could keep the memory of the spatial image and the accuracy was greater than in other methods (Hutchison and Loomis, 2006). The direct

action method is limited by three factors: 1) The initial visual coding of the target position, (2) Perception and integration of body movements by the observer over time, and (3) Motion control and response selection. The perceptual errors of observers often come from the perception and integration of the motion and the control and the selection of the motion. Since the direct action method involves the body movement of the subject, it takes a long time for each trial, and the subject and the experimenter have to spend a lot of time and energy.

4. Indirect action method

In the perceptual method and the direct action method, the target is in a straight line, so it can only prove that human beings can accurately perceive egocentric distance in a straight line. If observers have accurately perceived egocentric distance, they should be able to accurately represent the position of the target in more complex behavioral tasks such as multiple locations in space (Loomis et al., 1992). In the paper, the methods that require the observers to follow a curve (not a straight line) while walking or pointing toward the target with their eyes closed are called the Indirect Action Method, including triangulation-by-pointing and triangulation-by-walking. The core logic of the method is that after the stage of perceiving, observers do not directly walk towards the target, but walk along a path with a certain angle to the target. The method can help to confirm whether the representation of egocentric distance is limited by the straight line between the target and the observer.

4.1. Triangulation-by-pointing

Triangulation-by-pointing is also called continuous pointing. It refers to the measurement that after memorizing the location of the object, the individual closes his/her eyes, faces the object, and walks forward while continuously pointing in the direction of the remembered object (see Figure 8; Loomis et al., 1992, 1996; Fukusima et al., 1997; Burkitt et al., 2020). In the method, after viewing the target at the start point, the observer needs to stand flanker to the target. Then he/she must point with a finger in the direction of the target, and the experimenter will record the angle of the finger with an instrument. After that, the observer will close eyes and start to walk forward with pointing to the target. When the experimenter gives a sign, the observer will stop and the experimenter will record the angle of pointing. In addition to the way above, the experimenters can also record the direction of pointing, so that they can affirm whether the observer can represent the location of the target in different places. The perceptual distance in the method is indicated distance which can be calculated by some physical indicators, such as walking distance, the direction of pointing, and the distance between the target and the observer. To verify the effectiveness of triangulation-by-pointing, except for the closing eyes condition (without vision while walking), the

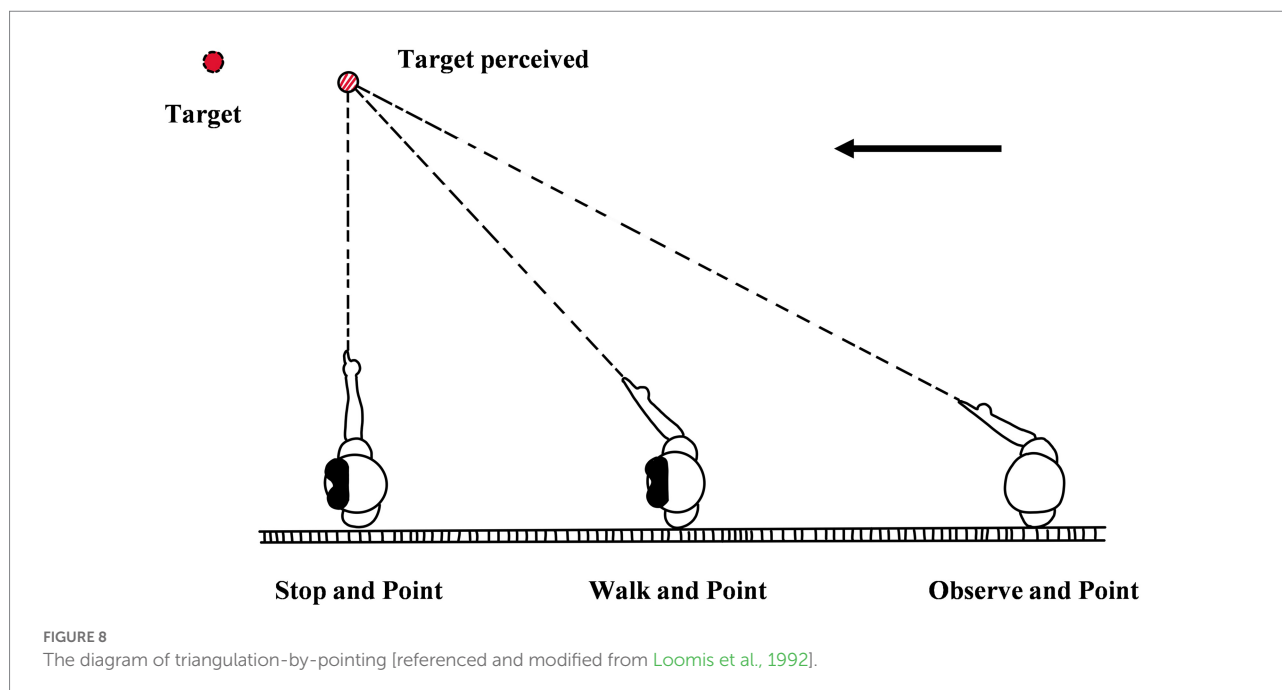
experimenters set another condition-opening eye condition (with vision input while walking) as a comparison; namely, the observer needs to keep eyes open while walking and pointing. The result shows that even in the condition without vision, the observer still can accurately perceive the distance up to 5.7 m. This means that the ability of human beings to perceive distance is not limited to the straight lines (Loomis et al., 1992, 1996; Loomis and Knapp, 2003), and the hypothesis that humans can accurately perceive egocentric distance is true (Fukusima et al., 1997).

The core logic of the method is that the observer does not need to walk directly to the target. Triangulation-by-pointing is similar to the situation in daily life in which an individual can bypass multiple obstacles to reach the bed in a dark room with the light just turned off. The experimental results of the method and the actual life experience both show that human beings can accurately perceive egocentric distance. The action in the method is similar to that in the directed action method, and they can both be classified as visually directed action. Due to this, it has the same advantages as the direct action method. Besides, the other advantages are as follows: (1) In the method, the position of the body, the relative position between the body and the target, and the pointing position keep updating, so the method involves rich psychological processes, such as path integration (Loomis and Knapp, 2003). (2) The method does not depend on the completeness of the path. Therefore, the method is suited to apply to an environment that involves puddles, water, and other obstacles which are not conducive to walking. (3) This method can be used in the study of egocentric distance perception involving location updating, so the researchers can investigate the observer's perception of the location of the object at any time, and then explore the observer's perception of egocentric distance in the whole space (rather than the narrow space between the observer and the object).

The limitation of the method is the small measurement scope. In other words, as the target becomes further away, the accuracy of measurement decreases significantly. Loomis et al. (1992) found that if the range of distance was beyond 5.7 m, the accuracy of the method would decrease rapidly (Loomis et al., 1992). Although the accurate distance range could be increased to 15 m by appending the walking distance of observers (Fukusima et al., 1997), the accuracy of the method was still lower than that of the direct action paradigm, such as blind walking (He et al., 2004; Ooi and He, 2007). In addition, the method has not overcome the methodological shortcomings of the perceptual method or the direct action method and the indicator of the method is hard to calculate, so this method is rarely used in current studies.

4.2. Triangulation-by-walking

Triangulation-by-walking refers to a method in which the observer, after remembering the position of the target, closes his/her eyes and walks along an oblique pathway to the target. After hearing the experimenter's command to stop, the observer

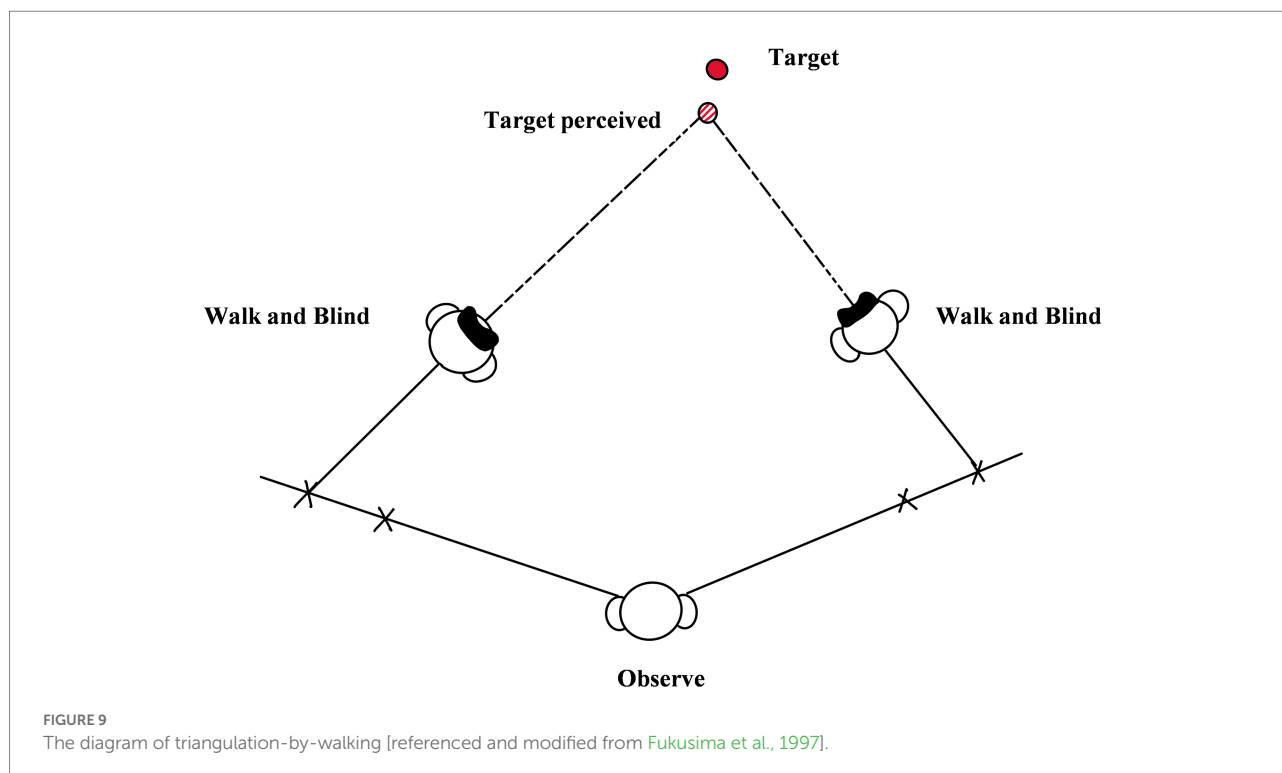


immediately stops and turns around to face the direction of the perceived target, and continues to walk (see Figure 9; Loomis et al., 1996; Fukusima et al., 1997; He et al., 2004; Li, 2017). Loomis et al. (1996) developed the measurement method and in the paper of Fukusima et al. (1997), they introduced the method in detail, involving core logic, measurement process, advantages, and disadvantages (Fukusima et al., 1997). As shown in Figure 9, in the method, the experimenter set up two oblique pathways of the same length on both sides of the starting point. The angle between the two pathways and the horizontal line was equal and the turning point was 4 meters and 6 meters. In each trial, the observer needed to stand at the starting point to observe the target, and after observing over, turn left or turn right to face a random pathway. And then, the observer needed to close eyes and walk along a straight-line slanting to the target. When the observer arrived at the turning point (4m or 6m), the experimenter would instruct him/her to turn around to the location of the target in mind with eyes closed. After turning around, the observer still needed to walk forward and try to arrive at the location of the target. When the observer stopped walking, the experimenter would record the direction of walking and then take the observer back to the starting point. Subsequently, the observer needed to go the other way and do the same things as in the first pathway. After the finishing of a trial, the experimenter could use the directions of starting point, turning point, and walking pathway after turning around to construct the terminal courses of the observer in the two pathways. The location of the target and the perceived distance can be calculated by the trigonometric function and the terminal courses. It is noted that there is no limitation in the number of oblique pathways, the position of turning, and the walking distance after turning.

The core logic, the advantages, and the disadvantages of triangulation-by-pointing and triangulation-by-walking are roughly the same. The difference between the two methods is that body turning is closer to human daily life and more natural than arm pointing, so the range of accurate measurement of triangulation-by-walking is greater than that of triangulation-by-pointing (Fukusima et al., 1997). And the triangulation-by-walking is suited to the measurement in the whole medium range of distance (2–25 m). However, when using the method in a narrow space, the bias will increase (Klein et al., 2009), so a large physical space is required.

4.3. The summary of indirect action method

Verifying the accuracy of distance perception in a more ecological and complex environment is a theoretical contribution of the Indirect Action Method and is also what inspired Loomis and Fukusima to develop the method (Loomis et al., 1992, 1996; Fukusima et al., 1997). From the perspective of cognitive psychology, egocentric distance perception is composed of several sub-psychological processes. The accuracy of egocentric distance perception also indicates the accuracy of the sub-psychological process, including: (1) perceiving accurately the location of the target, (2) perceiving accurately active self-motion, (3) in the process of moving, accurately imaginal updating the location of the target, (4) accurately pointing or walking to the updating position of the target (Loomis et al., 1992; Fukusima et al., 1997). What is noteworthy is that perceiving the location of the target accurately is not equal to constructing the clear spatial representation. The observer can both accurately perceive



egocentric distance in the condition with sufficient depth cues and in the condition without insufficient depth cues (Philbeck and Loomis, 1997).

The unique advantage of the method is that the method can be used to study spatial updating. Spatial (or imaginary) updating refers to the mental procedure in which the observer updates the relative position of the target and himself/herself by imagination while moving with eyes closed after remembering the initial position of the target (Loomis and Philbeck, 2008). The direct action method and the indirect action method both involve spatial updating, but the indirect action paradigm is the first method that combines spatial updating with trigonometric functions (Fukusima et al., 1997). In addition, the study on egocentric distance perception was further advanced by the method: the object of study is expanded from egocentric (or one-dimensional) distance to two-dimensional or even three-dimensional space, which improves ecological validity. In other words, the indirect action method measures not only the apparent distance but also the apparent location (Stefanucci and Proffitt, 2009). In the normal condition, when the target is on the ground, the observer can perceive the spatial updating through the reverse change between the angular declination below the horizon and the egocentric distance in the direct action method. However, if the angular declination below the horizon remains unchanged (e.g., the target is vertically upstairs or underfoot), the observer has to perceive the spatial updating by the change of egocentric distance, so in such condition, the indirect action method is more suitable than the direct action method. With the in-depth study, spatial updating has become a unique research topic in the field of spatial

perception, and the indirect action method has broad application prospects in the future.

5. Discussion

Distance perception is involved in many complex human behaviors, such as driving cars, flying airplanes, playing football, dancing, and so on. Understanding how to measure and similarities and differences between different methods will shed light on the causes of these complex behavior biases and further reduce human error in reality. According to the types of action involved in the measurement methods, nine methods of egocentric distance perception are divided into three categories, and the measurement steps, core logic, applicable scope, advantages and disadvantages of each are introduced, and the advancement of the inherent scientific issues is discussed. What noteworthy is that, although many methods have been used to measure distance perception (see Table 1), the scientific problem of how to measure or probe the psychological quantity of distance has not been satisfactorily solved.

It should be noted that the perception methods and action methods (both direct and indirect) may involve different psychological processes. In the perception methods, participants receive continuous visual input, so the measurement may be more relevant to the early perception stage (i.e., perception). However, in the action methods, the visual information is excluded in the measurement stage, which makes the action methods may rely on the distance representation constructed in the observer's memory

TABLE 1 The summary of advantages and disadvantages of each method for measuring distance perception.

Methods	Advantages	Disadvantages
Verbal report	Intuitively clarify and simple	Influenced by cognitive correction and the observer's cognition of distance; not suitable to be used alone
Perceptual distance matching	accurate; be applied in the vista space (> 30 m) and used in both real environment and VR environment	Affected by participant's cognitive factors
Successive Equal-appearing Intervals of Distance Judgment Measurement		Confuse the measurements of exocentric distance and egocentric distance
Blind Walking	Accurate	Take time and effort; easily produce practice and fatigue effects; difficultly applied to far space
Blind-walking gesturing	Integrate the distance and direction judgment, accurate	The same as blind walking
Blindfolded throwing	Simple and time-saving	Participants are required to have a high level of sports ability; inaccurate
Blind rope pulling	Accurate, can be applied to study the elder and patients, even in a narrow space	Participants need to be familiar with rope pulling. The complexity of experimental design affects task performance.
Triangulation-by-pointing	Investigate the rich psychological process; expand the research topic to 3D space	Inaccurate; complex calculation procedure, require larger test space, and takes time and effort
Triangulation-by-walking	More accurate than triangulation-by-pointing, other advantages are the same as Triangulation-by-pointing	The same as Triangulation-by-pointing

(i.e., spatial representation). The relationship between these two psychological processes (completely different or both belong to the perception) still needs to be explored in the future. In the actual study, researchers will choose the appropriate experimental method according to research topics, research objects, and experimental situations. Perceptual distance matching, blind walking, and blind-walking gesturing have higher accuracy; the application scope of perceptual distance matching is large; verbal report is easy to understand and has low space requirement; blindfolded throwing and blind rope pulling are suitable for the larger range of people; triangulation-by-pointing and triangulation-by-walking extend the study object to the whole space. According to the study topic, perceptual distance matching can not only be used to study egocentric distance in the real world but also be used to study egocentric distance and exocentric distance in virtual reality (Loomis et al., 1996; Ooi and He, 2007). For the subjects, all three categories can be applied to normal healthy subjects, but to the special subjects who are difficult to walk, such as elders, stroke patients, and disabled people, the better methods are the perceptual method, blindfolded throwing, and blind rope pulling (Philbeck et al., 2010; Bian and Andersen, 2013). Egocentric distance perception involves three environments: the real world, the virtual world with HMD (head monitor display), and the virtual world with the big screen on the wall. The change in the research environment determines the choice of method. For example, timed imagined walking (a variant of blind walking) is suitable to the environment with the big screen on the wall, blind treadmill walking, blind walking, triangulation-by-pointing, and triangulation-by-walking are suitable to the virtual world with HMD. More recently, a growing number of researchers have combined multiple methods to further improve measurement accuracy based on the convergence

principle (Tenhundfeld and Witt, 2017; Rand et al., 2019). Such as combining the verbal report with blind walking (Philbeck and Loomis, 1997), combining blind throwing, triangulation-by-walking, and blind walking to explore the frame of the ground surface and verify the SSIP (He et al., 2004). However, the problems such as what the standard of the convergence principle is and how to deal with the conflicting measurement results under the convergence principle still need to be further solved.

There is no doubt that the method promotes the field of egocentric distance perception, but there are still some common problems that need to be further discussed. The first is the inconsistencies between the results of different methods (Renner et al., 2013). For example, the result of the direct action method is larger than that of verbal report (Philbeck and Loomis, 1997), and the variability of the indirect action method is larger than that of the direct action method (Loomis et al., 1992). The perceptual method is mediated by consciousness and shows the perceptual results through language or matching. The behavioral method (including direct and indirect) bypasses perception and directly shows perceptual results through action. The neural mechanisms of perception and behavioral control are very different: The is mainly controlled by the ventral pathway, while the latter is mainly determined by the dorsal pathway (Loomis and Philbeck, 2008). The difference between methods may reflect the existence of multiple neural structures for distance perception in the human brain. The problem of “what are the differences between methods of egocentric distance perception” may be transformed into “what are the activation conditions of these subcategories”, “what are the mechanisms and rules of subcategories of egocentric distance perception” and so on. The second is about the size of the target. In the experiments of egocentric distance, multiple distance conditions are usually set up simultaneously to investigate the

perceptual rule of human beings in the whole space. At this time, the size of the target at each distance becomes a problem that cannot be avoided, because of the covariance of magnitude and distance (Gilinsky, 1951; Loomis and Philbeck, 2008). There are two common approaches to the problem. The one is to keep the physical size of the target; for example, place foam balls 10 cm in diameter at each distance. This method has high ecological validity, but with the increase of the distance, the size of the target in the retina (a kind of depth cue) gradually decreases. When the study problems are about other depth cues (such as binocular parallax or texture gradient), retinal image size (or view angle) will become an interfering variable. Because when the observer views the target, he/she can judge egocentric distance by the size of the retinal image alone and the size perception may replace the distance perception. The two is to keep the retinal image size (or view angle) of the target. Such as, foam balls with a diameter of 1-degree view angle are used at each distance. With the increase of distance, the physical size of the target will gradually increase. Although the method eliminates the interference of view angle, at this time, subjects perceive different objects at different distances. So, it is difficult to compare experimental effects at different distances. Whether keeping physical size or retinal size is a practical problem faced by most spatial perception researchers, the optimal solution has not been formed yet. Researchers can only make choices according to their interested problems. The three is that the methods of static egocentric distance perception measure the perceptual results at the behavioral level and it is difficult to directly measure the perception. Since the internal mechanism of different methods is different, it needs to be further discussed whether the difference comes from internal generative mechanism. The four is that the mechanism and characteristics of spatial perception are mainly discussed in the psychological research, but the relationship between spatial perception and daily life or production still needs to be further strengthened. In the future, psychological studies of space may revolve around the topics of greater application value (e.g., environmental psychology).

With the development of the Metaverse, VR and AR technologies are developing rapidly. The development of these two technologies relies on the research of egocentric distance perception. As far as VR technology is concerned, users should experience a virtual environment that is like our real world to achieve high immersion. For AR technology, users need a high degree of integration of virtual information and the real environment. However, the problem of distance compression has been hindering the VR experience (Kelly et al., 2018) and AR experience (Singh et al., 2018). As a result, VR scenes cannot provide authentic high immersion, and AR content cannot be accurately superimposed on reality. The study of egocentric distance perception can play a role in solving the problem of distance compression. First, it provides a theoretical basis for the cause of distance compression. Second, it provides behavioral indicators for distance compression to measure the degree of compression and verify the effectiveness of solutions. From the perspective of practical needs, how to build a virtual world

identical to the real world (Virtual Reality technology) has become an important issue concerned by the government, academia, and even the business community. In the process of product and technology development, developers often use a static method to verify the effect of various factors on egocentric distance perception in VR. For example, the effects of peripheral visual field stimulus presentation, visual field size in virtual space, and other factors on egocentric distance perception were examined using the direct action method (Vaziri et al., 2021), the effects of video card quality, mechanical properties of a head-mounted display, binocular parallax, head scaling and sound scaling on egocentric distance perception was verified by the indirect action method (Choudhary et al., 2021). Although some studies have proved that there is no significant difference in the accuracy of egocentric distance perception between blindfolded throwing and blind walking method in real or virtual environments (Sahm et al., 2005), the construction of realistic depth sense is still the core focus and difficulty in VR (Loomis and Knapp, 2003). The premise to solve the problem is to find or develop a method that can accurately measure depth perception (i.e., egocentric distance perception), that is to find the optimal or most appropriate egocentric distance perception method.

Although this paper focuses on egocentric distance perception in visual modality, auditory distance is a very important research topic. On the one hand, although the current research focuses on egocentric distance in the visual modality, most researchers did not eliminate auditory clues or recruit blind people in the experiments. In other words, there may be an effect of auditory distance perception in the methods mentioned in this paper. It is worth noting that for blind people, auditory distance perception is more important (Voss et al., 2004; Kolarik et al., 2013a,b; Cappagli et al., 2017; Cuturi et al., 2021). In these papers, researchers used different methods (e.g., verbal reporting vs. direct action such as triangle completion) to measure egocentric distance perception. It is also necessary to review this research field. In the future, researchers may focus on this topic.

Author contributions

BD and XT had the idea for the article. AC and XZ performed the literature research. BD, YS and ZG wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (32100841, 32100842, and 71974140), the MOE Project of Humanities and Social Sciences Grant (20YJC190002), the Project of Social Science Foundation of Jiangsu Province (20YJC008), Jiangsu University Philosophy and

Social Science Research Project (2019SJA1267), and the Qing Lan Project of Jiangsu Universities.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Bernhard, C., Reinhard, R., Kleer, M., and Hecht, H. (2021). A case for raising the camera: a driving simulator test of camera-monitor systems. *Hum. Factors*. 001872082110109. doi: 10.1177/00187208211010941
- Bian, Z., and Andersen, G. J. (2013). Aging and the perception of egocentric distance. *Psychol. Aging* 28, 813–825. doi: 10.1037/a0030991
- Blake, R., and Sekuler, R. (2006). "Seeing a three-dimensional world," in *Perception*. 5th Edn. eds. R. Blake and R. Sekuler (New York: McGraw-Hill Higher Education), 273–315.
- Bossard, M., Goulon, C., and Mestre, D. (2020). Viewpoint oscillation improves the perception of distance travelled in static observers but not during treadmill walking. *Exp. Brain Res.* 238, 1073–1083.
- Bufacchi, R. J., and Iannetti, G. D. (2018). An action field theory of Peripersonal space. *Trends Cogn. Sci.* 22, 1076–1090. doi: 10.1016/j.tics.2018.09.004
- Burkitt, J. J., Campos, J. L., and Lyons, J. L. (2020). Iterative spatial updating during forward linear walking revealed using a continuous pointing task. *J. Mot. Behav.* 52, 145–166. doi: 10.1080/00222895.2019.1599807
- Cappagli, G., Cocchi, E., and Gori, M. (2017). Auditory and proprioceptive spatial impairments in blind children and adults. *Dev. Sci.* 20:e12374. doi: 10.1111/desc.12374
- Cavallo, V., Colomb, M., and Doré, J. (2001). Distance perception of vehicle rear lights in fog. *Hum. Factors* 43, 442–451. doi: 10.1518/001872001775898197
- Chen, N., Cai, P., Zhou, T., Thompson, B., and Fang, F. (2016). Perceptual learning modifies the functional specializations of visual cortical areas. *Proc. Natl. Acad. Sci.* 113, 5724–5729. doi: 10.1073/pnas.1524160113
- Choudhary, Z., Gottsacker, M., Kim, K., Schubert, R., Stefanucci, J., Bruder, G., et al. (2021). "Revisiting distance perception with scaled embodied cues in social virtual reality", in: 2021 IEEE Virtual Reality and 3D User Interfaces (VR): IEEE), 788–797.
- Cutting, J. E. (1995). "Perceiving layout and knowing distances : the integration, relative potency, and contextual use of different information about depth," in *Handbook of Perception and Cognition*. 69–117.
- Cuturi, L. F., Alborn, P., Cappagli, G., Volta, E., Volpe, G., and Gori, M. (2021). The influence of yaw rotation on spatial navigation during development. *Neuropsychologia* 154:107774. doi: 10.1016/j.neuropsychologia.2021.107774
- Dong, B., Chen, A., Zhang, T., and Zhang, M. (2021a). Egocentric distance perception disorder in amblyopia. *Psychol. Belgica* 61, 173–185. doi: 10.5334/pb.1038
- Dong, B., Chen, A., Zhang, Y., Zhang, Y., Zhang, M., and Zhang, T. (2021b). The foggy effect of egocentric distance in a nonverbal paradigm. *Sci. Rep.* 11:14398. doi: 10.1038/s41598-021-93380-9
- Eby, D. W., and Loomis, J. M. (1987). A study of visually directed throwing in the presence of multiple distance cues. *Percept. Psychophys.* 41, 308–312. doi: 10.3758/bf03208231
- Fukushima, S. S., Loomis, J. M., and Da Silva, J. A. (1997). Visual perception of egocentric distance as assessed by triangulation. *J. Exp. Psychol. Hum. Percept. Perform.* 23, 86–100. doi: 10.1037/0096-1523.23.1.86
- Gibson, J. J. (1950). *The Perception of the Visual World*. Cambridge, MA: The Riverside Press.
- Gilinsky, A. S. (1951). Perceived size and distance in visual space. *Psychol. Rev.* 58, 460–482. doi: 10.1037/h0061505
- Hagen, M. A., and Teghtsoonian, M. (1981). The effects of binocular and motion-generated information on the perception of depth and height. *Percept. Psychophys.* 30, 257–265. doi: 10.3758/BF03214281
- Hajnal, A., Bunch, D. A., and Kelty-Stephen, D. G. (2016). Pulling out all the stops to make the distance: effects of effort and optical information in distance perception responses made by rope pulling. *Atten. Percept. Psychophys.* 78, 685–699. doi: 10.3758/s13414-015-1035-x
- He, Z. J., Wu, B., Ooi, T. L., Yarbrough, G., and Wu, J. (2004). Judging egocentric distance on the ground: occlusion and surface integration. *Perception* 33, 789–806. doi: 10.1068/p5256a
- Hutchison, J., and Loomis, J. (2006). Does energy expenditure affect the perception of egocentric distance? A failure to replicate experiment 1 of Proffitt, Stefanucci, Banton, and Epstein (2003). *Span. J. Psychol.* 9, 332–339. doi: 10.1017/S1138741600006235
- Kelly, J. W., Cherep, L. A., Klesel, B., Siegel, Z. D., and George, S. (2018). Comparison of two methods for improving distance perception in virtual reality. *Acm transactions on applied. Perception* 15, 1–11. doi: 10.1145/3165285
- Klein, E., Swan, J., Schmidt, G., Livingston, M., and Staadt, O. (2009). Measurement protocols for medium-field distance perception in large-screen immersive displays, in: 2009 IEEE Virtual Reality Conference. (Lafayette, LA, USA: IEEE).
- Kolarik, A. J., Cirstea, S., and Pardhan, S. (2013a). Evidence for enhanced discrimination of virtual auditory distance among blind listeners using level and direct-to-reverberant cues. *Exp. Brain Res.* 224, 623–633. doi: 10.1007/s00221-012-3340-0
- Kolarik, A. J., Cirstea, S., Pardhan, S., and Moore, B. C. (2013b). An assessment of virtual auditory distance judgments among blind and sighted listeners. Paper presented at the Proceedings of Meetings on Acoustics Montreal, 19.
- Li, B. (2017). Distance perception in virtual environment through head-mounted displays. Unpublished Doctoral Dissertation, Michigan Technological University.
- Loomis, J. M., Da Silva, J. A., Philbeck, J. W., and Fukushima, S. S. (1996). Visual perception of location and distance. *Curr. Dir. Psychol. Sci.* 5, 72–77. doi: 10.1111/1467-8721.ep10772783
- Loomis, J. M., and Knapp, J. (2003). "Visual perception of egocentric distance in real and virtual environments" in *Virtual and Adaptive Environments: Applications, Implications, and Human Performance Issues*. eds. L. J. Hettinger and M. W. Haas (Mahwah, NJ: Lawrence Erlbaum Associates Publishers).
- Loomis, J. M., and Philbeck, J. W. (2008). "Measuring spatial perception with spatial updating and action" in *Carnegie Mellon Symposia on Cognition. Embodiment, Ego-space, and Action*. eds. R. L. Klatzky, B. MacWhinney and M. Behrman (New York, United States: Psychology Press), 1–43.
- Loomis, J. M., Silva, J. A. D., Fujita, N., and Fukushima, S. S. (1992). Visual space perception and visually directed action. *J. Exp. Psychol. Hum. Percept. Perform.* 18, 906–921.
- Mershon, D. H., Kennedy, M., and Falacara, G. (1977). On the use of 'calibration equations' in perception. *Research* 6, 299–311. doi: 10.1068/p060299
- Napieralski, P. E., Altenhoff, B. M., Bertrand, J. W., Long, L. O., Babu, S. V., Pagano, C. C., et al. (2011). Near-field distance perception in real and virtual environments using both verbal and action responses. *ACM Trans. Appl. Percept.* 8, 1–19. doi: 10.1145/2010325.2010328
- Ooi, T. L., and He, Z. J. (2007). A distance judgment function based on space perception mechanisms: revisiting Gilinsky's (1951) equation. *Psychol. Rev.* 114, 441–454. doi: 10.1037/0033-295X.114.2.441
- Ooi, T. L., Wu, B., and He, Z. J. (2001). Distance determined by the angular declination below the horizon. *Nature* 414, 197–200. doi: 10.1038/35102562
- Ooi, T. L., Wu, B., and He, Z. J. (2006). Perceptual space in the dark affected by the intrinsic bias of the visual system. *Perception* 35, 605–624. doi: 10.1068/p5492
- Peer, A., and Ponto, K. (2016). "Perceptual space warping: preliminary exploration" in 2016 IEEE virtual reality (VR) (SC, USA: Greenville)

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Philbeck, J. W., and Loomis, J. M. (1997). Comparison of two indicators of perceived egocentric distance under full-cue and reduced-cue conditions. *J. Exp. Psychol. Hum. Percept. Perform.* 23, 72–85.
- Philbeck, J. W., Woods, A. J., Kontra, C., and Zdenkova, P. (2010). A comparison of blindpulling and blindwalking as measures of perceived absolute distance. *Behav. Res. Methods* 42, 148–160. doi: 10.3758/brm.42.1.148
- Plumert, J. M., Kearney, J. K., Cremer, J. F., and Recker, K. (2005). Distance perception in real and virtual environments. *ACM Trans. Appl. Percept.* 2, 216–233. doi: 10.1145/1077399.1077402
- Proffitt, D. R. (2006). Distance perception. *Curr. Direct. Psychol. Res.* 15, 131–135. doi: 10.1111/j.0963-7214.2006.00422.x
- Rand, K. M., Barhorst-Cates, E. M., Kiris, E., Thompson, W. B., and Creem-Regehr, S. H. (2019). Going the distance and beyond: simulated low vision increases perception of distance traveled during locomotion. *Psychol. Res.* 83, 1349–1362. doi: 10.1007/s00426-018-1019-2
- Renner, R. S., Velichkovsky, B. M., and Helmert, J. R. (2013). The perception of egocentric distances in virtual environments - a review. *ACM Comput. Surv.* 46, 1–40. doi: 10.1145/2543581.2543590
- Sahm, C. S., Creem-Regehr, S. H., Thompson, W. B., and Willemsen, P. (2005). Throwing versus walking as indicators of distance perception in similar real and virtual environments. *ACM Trans. Appl. Percept.* 2, 35–45. doi: 10.1145/1048687.1048690
- Santillán, J. E., and Barraza, J. F. (2019). Distance perception during self-movement. *Hum. Mov. Sci.* 67:102496. doi: 10.1016/j.humov.2019.102496
- Siegel, Z. D., and Kelly, J. W. (2017). Walking through a virtual environment improves perceived size within and beyond the walked space. *Atten. Percept. Psychophys.* 79, 39–44.
- Sinai, M. J., Ooi, T. L., and He, Z. J. (1998). Terrain influences the accurate judgement of distance. *Nature* 395, 497–500. doi: 10.1038/26747
- Singh, G., Ellis, S. R., and Swan, J. E. (2018). The effect of focal distance, age, and brightness on near-field augmented reality depth matching. *IEEE Trans. Vis. Comput. Graph.* 26, 1385–1398. doi: 10.1109/TVCG.2018.2869729
- Stefanucci, J. K., and Proffitt, D. R. (2009). The roles of altitude and fear in the perception of height. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 424–438. doi: 10.1037/a0013894
- Teng, D. W., Eddy, C. L., and Kelty-Stephen, D. G. (2016). Non-visually-guided distance perception depends on matching torso fluctuations between training and test. *Atten. Percept. Psychophys.* 78, 2320–2328.
- Tenhundfeld, N. L., and Witt, J. K. (2017). Distances on hills look farther than distances on flat ground: evidence from converging measures. *Atten. Percept. Psychophys.* 79, 1165–1181. doi: 10.3758/s13414-017-1305-x
- Thomson, J. A. (1983). Is continuous visual monitoring necessary in visually guided locomotion? *J. Exp. Psychol. Hum. Percept. Perform.* 9, 427–443. doi: 10.1039/c4lc00876f
- Vaziri, K., Bondy, M., Bui, A., and Interrante, V. (2021). "Egocentric distance judgments in full-Cue video-see-through VR conditions are no better than distance judgments to targets in a void", in: 2021 IEEE Virtual Reality and 3D User Interfaces (VR). (Lisboa, Portugal: IEE).
- Voss, P., Lassonde, M., Gougoux, F., Fortin, M., Guillemot, J. P., and Lepore, F. (2004). Early- and late-onset blind individuals show supra-normal auditory abilities in far-space. *Curr. Biol.* 14, 1734–1738. doi: 10.1016/j.cub.2004.09.051
- Wallin, C. P., Gajewski, D. A., Teplitz, R. W., Mihelic Jaidzeka, S., and Philbeck, J. W. (2017). The roles for prior visual experience and age on the extraction of egocentric distance. *J. Gerontology: Series B* 72, 91–99.
- Wu, B., He, Z. J., and Ooi, T. L. (2007a). Inaccurate representation of the ground surface beyond a texture boundary. *Perception* 36, 703–721. doi: 10.1068/p5693
- Wu, B., He, Z. J., and Ooi, T. L. (2007b). The linear perspective information in ground surface representation and distance judgment. *Percept. Psychophys.* 69, 654–672. doi: 10.3758/BF03193769
- Zhou, L., Deng, C., Ooi, T. L., and He, Z. J. (2016). Attention modulates perception of visual space. *Nat. Hum. Behav.* 1. doi: 10.1038/s41562-016-0004



OPEN ACCESS

EDITED BY

Alyssa A. Brewer,
University of California,
Irvine, United States

REVIEWED BY

Charlie Frowd,
University of Central Lancashire,
United Kingdom
Po-Hung Lin,
National Kaohsiung University of
Science and Technology,
Taiwan
Ding-Hau Huang,
National Taipei University of Business,
Taiwan

*CORRESPONDENCE

Yi-Lang Chen
✉ ylchen@mail.mcut.edu.tw

SPECIALTY SECTION

This article was submitted to
Perception Science,
a section of the journal
Frontiers in Psychology

RECEIVED 28 October 2022

ACCEPTED 03 January 2023

PUBLISHED 17 January 2023

CITATION

Chen Y-L, Wu C-Y, Li S-C, Yu T-M and Yu S-P
(2023) Effect of mask coverage on face
identification in Taiwanese men and women.
Front. Psychol. 14:1082376.
doi: 10.3389/fpsyg.2023.1082376

COPYRIGHT

© 2023 Chen, Wu, Li, Yu and Yu. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License](#)
(CC BY). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted which
does not comply with these terms.

Effect of mask coverage on face identification in Taiwanese men and women

Yi-Lang Chen*, Cheng-Yu Wu, Shih-Cheng Li, Tai-Min Yu and
Shu-Ping Yu

Department of Industrial Engineering and Management, Ming Chi University of Technology, New Taipei, Taiwan

Mask wearing is the easiest and most effective way to avoid COVID-19 infection; however, it affects interpersonal activities, especially face identification. This study examined the effects of three mask coverage levels (full coverage, FC; coverage up to the middle [MB] or bottom of the nose bridge [BB]) on face identification accuracy and time. A total of 115 university students (60 men and 55 women) were recruited to conduct a computer-based simulation test consisting of 30 questions (10 questions [five face images each of men and women] for the three mask coverage levels). One unmasked target face and four face images with a specified mask coverage level were designed for each question, and the participants were requested to select the same face from the four covered face images on the basis of the target face. The ANOVA results indicated that identification accuracy was significantly affected by sex ($p < 0.01$) and the mask coverage level ($p < 0.001$), whereas identification time was only influenced by sex ($p < 0.05$). The multiple comparison results indicated that the identification accuracy rate for faces wearing a mask with FC (90.3%) was significantly lower than for those wearing masks with coverage up to the MB (93.7%) and BB (94.9%) positions; however, no difference in identification accuracy rate was observed between the MB and BB levels. Women exhibited a higher identification accuracy rate than men (94.1% vs. 91.9%) in identifying unfamiliar faces, even though they may spend less time identifying the images. A smaller mask coverage level (i.e., the BB level) does not facilitate face identification. The findings can be served as a reference for people to trade-off between wearing a mask and interpersonal interaction in their daily activities.

KEYWORDS

mask coverage, face identification, identification accuracy, sex, identification time

Introduction

Since the outbreak of the COVID-19 pandemic in 2020, more than 500 million people have been infected, resulting in more than 6.3 million deaths worldwide (World Health Organization [WHO], 2022). Mask wearing is the easiest and most effective method to avoid coronavirus infection. The United States Centers for Disease Control and Prevention (CDC, 2020) has suggested that all people should cover their lower face in public settings. Governments around the world have also made similar suggestions, often requiring citizens to wear face masks in public areas or on public transport (Al Jazeera News, 2020). In Taiwan, without exception, people had been stipulated to wear masks in public spaces from December 2020, and the regulations continue to this day (Taiwan Centers for Disease Control [TCDC], 2022). Although the CDC has recommended a standard protocol for wearing masks, however, people alter the coverage of face masks to maximize their comfort level or to facilitate identification by others.

Although masks are effective in inhibiting COVID-19 spread, their implementation has caused changes in people's lifestyles. People's facial expressions are covered when using a face mask, thus affects their feelings and cognitions during interpersonal interactions (Cartaud et al., 2020). Wearing masks in public poses a challenge for requiring facial recognition and recognition. A study reported crimes committed by individuals wearing face masks, presumably to disguise or hide their appearance (Southall and Van Syckle, 2020). Faces are often used as a way of verifying identity of an individual, whether across borders or buying alcohol at the local stores, but masks hinder these recognitions (Carragher and Hancock, 2020). Mask wearing also affects interpersonal activities, for example, individuals may have difficulty in identifying known people. Although the identification of people with partially covered or occluded faces is challenging, relatively little research has focused on how covering the lower face influences perceptual face identifications.

Because face masks conceal the lower face (i.e., the mouth and nose), they hinder social interactions and identifications. Recent studies have examined the effects of face masks on face-matching performance (Carragher and Hancock, 2020; Noyes et al., 2021). Although large differences exist among individuals (Estudillo et al., 2021), face masks impair overall face-matching performance (Carragher and Hancock, 2020; Noyes et al., 2021). Factors affecting face identification accuracy have also been explored. For example, an unfamiliar face matching is often error-prone under optimal conditions (Burton et al., 2010; Kramer et al., 2018), and performance deteriorates further in worse conditions (Fysh and Bindemann, 2017). Even minor differences between the images can influence the accuracy, such as color images or black and white images (Bobak et al., 2019), the distances between the individual and the camera in each image (Noyes and Jenkins, 2017), the image quality (Bindemann et al., 2013), the shown viewpoints of the faces (Estudillo and Bindemann, 2014), spectacle usage (Graham and Ritchie, 2019; Noyes et al., 2021), and the lighting conditions (Hill and Bruce, 1996). Moreover, a long time interval between the identification of two images may reduce the accuracy (Megreya et al., 2013). Even if the two pictures are taken just minutes apart, participants make errors in approximately 20% of trials (Burton et al., 2010). Difficulty in unfamiliar face matching was observed for the comparisons of two images (Megreya and Burton, 2006; Burton et al., 2010) and an image with a real person (White et al., 2014). By contrast, Carragher and Hancock (2020) discovered that masks have a large damaging effect on face-matching performance, and the degree of impairment was similar in the identification of both familiar and unfamiliar faces. Noyes et al. (2021) suggested that occlusion impedes face identification accuracy irrespective of a familiar or unfamiliar target. Therefore, masks complicate the face identification process, thereby warranting further exploration.

Most previous behavioral studies concerning sex effect on face identification took use of a learning-test paradigm (McKelvie, 1976; Lewin and Herlitz, 2002; Herlitz and Rehnman, 2008; Megreya et al., 2011). During a typical test, a serial list of unfamiliar faces was displayed for few seconds each, and participants were requested to memorize these faces, as they had to identify them from unfamiliar faces later. In such studies, women superiority in face identification was consistently represented by their higher accuracy compared to men (Megreya et al., 2011; Godard et al., 2013; Hansen et al., 2021; Wong and Estudillo, 2022). Furthermore, Sun et al. (2017) used a modified delayed matching-to-sample task to investigate the time course characteristics of face identification by event-related potential (ERP) for both sexes. This quantitative study using ERP

technique also verified that women were more accurate and faster than men on the task. However, the effect of mask coverage on face identification for both sexes was not examined in these investigations. The issue is even more relevant during the COVID-19 pandemic.

Several functions are impeded by wearing masks (e.g., speech, respiration, and comfort; Kumar and Lee, 2020; Rahne et al., 2021; Zhang et al., 2022); therefore, the compound effect of face masks could have a relevant impact on daily life communication even in those with normal respiration. Although the CDC has recommended a standard protocol for wearing masks, we observed that people alter the coverage of face masks (Figure 1) to maximize their comfort level or to facilitate identification by others. According to a survey of Indonesian residents, only 34.3% of the subjects wore face masks properly (Siahaan et al., 2021). Ganczak et al. (2021) further discovered that uncovered noses were the most frequent incorrect practices for masks wearing, with approximately 50% in the country. The COVID-19 pandemic has initiated several studies on face masks; however, no study has evaluated the effect of different mask coverage levels on face identification. Therefore, this study recruited 115 men and women as participants to examine face identification accuracy and time for three mask coverage levels (FC, full coverage; MB, coverage up to the middle nose bridge; and BB, coverage up to the bottom nose bridge; Figure 1). This study hypothesized that a decrease in mask coverage level would increase face identification accuracy.

Materials and methods

Participants

A total of 115 university students (60 men and 55 women) with a mean (standard deviation) age of 21.6 (2.8) and 22.8 (5.2) years for men and women, respectively, were recruited. All participants had familiarity with computers (which was required for the test). No participant exhibited self-reported vision defects, such as color blindness or color weakness, which were also checked by an on-line test.¹ Participants with no vision defects with the naked eye or after vision correction (i.e., wearing glasses) were included. All participants provided informed consent before participating in the study, and the study was approved by the Ethics Committee of Chang Gung Memorial Hospital, Taiwan.

Stimuli

To examine the effect of mask coverage levels on face identification, three coverage levels (FC, MB, and BB) were adopted. The face images for the test were obtained from the Asian Face Age Dataset (Niu et al., 2016). The screening of these face images was mainly based on the absence of special expressions of emotion to minimize the influence in face identification. We randomly selected 40 frontal male and female face images (20 each); of these, 10 were selected as target images (5 men and 5 women). The images were edited using Adobe Illustrator and Adobe Photoshop 23.5 (Adobe Systems, San Jose, CA, United States) and converted black and white, and the exposure was also standardized. To avoid the image resolution affecting the accuracy of the test, the resolution displayed on a 40-in screen (L42-6500, BenQ, Taipei, Taiwan)

¹ <https://www.ifreesite.com/color-blindness>

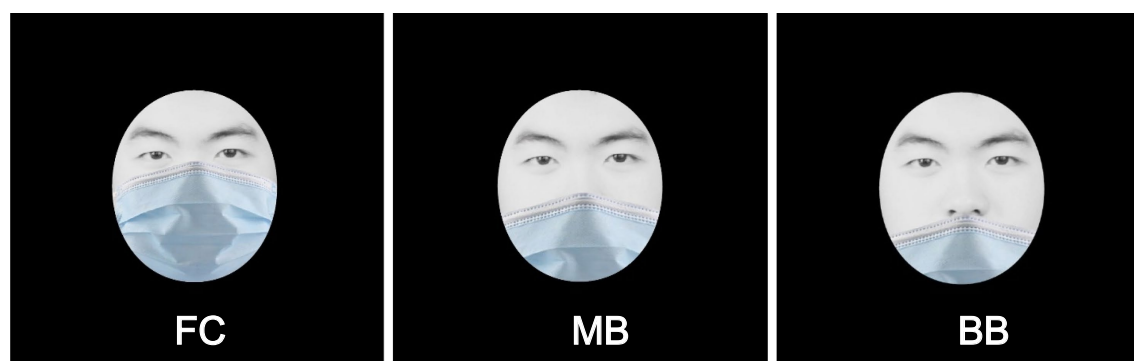


FIGURE 1

Three common mask coverage levels (FC, full coverage; MB, coverage up to the middle nose bridge; and BB, coverage up to the bottom nose bridge).

was controlled at $1,920 \times 1,080$ pixels. The plane medical mask (W17.5 cm \times H 9.5 cm, Vicvin Tech., Hsinchu, Taiwan) was then superimposed on the faces of the original image to achieve the three coverage levels. A total of 10 questions were developed for each coverage level, and the task involved a target face with a face covering and four choices presented on the subsequent page. In addition to the 10 target images, 30 images were used as non-answer options.

Experimental design and procedure

Because 10 questions (corresponding to 5 men and 5 women) were developed for each coverage level, the test contained 30 questions for the three levels. A total of 3,450 data sets (115 participants \times 3 coverage levels \times 10 questions) were recorded, including the participants' answers and the corresponding identification time. The arrangement of the testing combinations was completely randomized to avoid learning or cumulative bias.

The test was conducted in an isolated quiet room. The participants were requested to sit and follow the instructions displayed on the screen before proceeding to the test. Participants were allowed to adjust the seat height and keyboard position to ensure that their sights were perpendicular to the screen and they were comfortable throughout the experiment. A two-stage practice was conducted before data collection. The first stage was conducted to familiarize the participants with the usage of a computer keyboard and mouse, whereas the second stage was a pilot test with a similar question format to the final test (Figure 2). After that, an alert box was displayed to inform the participants that the practice was complete, and the formal test was going to begin. Upon pressing the OK button, the screen displayed a 8-s countdown to provide time for the participants to answer each question. Once the time has passed and the participant had not answered the question, it was regarded as invalid data, and the screen automatically jumped to the next question. In the test, the participants were requested to key in one correct answer (from four options) for each question until all 30 questions were completed. The total testing time for each participant was approximately 20 min, including instruction, preparation, and formal test.

Statistical analysis

Statistical analyses were conducted using SPSS 22.0, with significance defined as a minimum α level of 0.05 for all tests. Data

collected from the participants were analyzed using descriptive statistics (i.e., means and standard deviations). Two-way analysis of variance (ANOVA) was used to examine the effects of the three coverage levels (FC, MB, and BB) and sex on face identification accuracy and time, and Tukey's honestly significant difference test was used for multiple comparisons. In addition, one-way ANOVA was conducted to examine the effect of the mask coverage levels on the responses of each sex. A power value was used to examine if the effect size of any significant independent variable was satisfactory (i.e., power ≥ 0.8) as suggested by Cohen (1988).

Results

Two-way ANOVA

The two-way ANOVA results of face identification accuracy and time are listed in Tables 1, 2, respectively. Both sex ($p < 0.01$) and mask coverage levels ($p < 0.001$) significantly affected face identification accuracy. Although sex variable affected identification time ($p < 0.05$), the power value was less than 0.8. Women exhibited a higher accuracy rate (94.1% vs. 91.9%) and shorter identification time (2.60 vs. 2.78 s) than men. No interaction was observed in the ANOVA results. Tukey's test (Table 3) demonstrated that face identification accuracy was lower in the FC level than in the MB and BB levels. No difference in face identification accuracy was observed between the non-FC levels.

One-way ANOVA results for men and women

The effects of mask coverage levels on the two responses of men and women as analyzed using one-way ANOVA are listed in Table 4. The mask coverage levels significantly affected face identification accuracy but not identification time, regardless of sex. However, men exhibited a significant difference in face identification accuracy between the FC and BB levels. Women exhibited no difference in face identification accuracy between the MB and BB levels, with accuracy in these levels being different from that in the FC level. Figure 3 illustrates the six combinations formed using sex and coverage variables. The trend of face identification accuracy in the MB level is different between sexes.



FIGURE 2
A pilot test format similar to the final test with a full coverage level.

TABLE 1 Two-way ANOVA results of face identification accuracy.

Sources	SS	df	MS	F	value of p	Power
Sex	0.042	1	0.042	8.535	<0.01	0.829
Mask coverage	0.129	2	0.065	13.270	<0.001	0.998
Sex \times mask coverage	0.006	2	0.003	0.592	0.554	0.149

TABLE 2 Two-way ANOVA results of identification time.

Sources	SS	df	MS	F	value of p	Power
Sex	2.606	1	2.606	4.752	<0.05	0.585
Mask coverage	2.656	2	1.328	2.422	0.090	0.487
Sex \times mask coverage	0.114	2	0.057	0.104	0.902	0.066

Discussion

Although studies have indicated that mask wearing affects face identification accuracy (Carragher and Hancock, 2020; Noyes et al., 2021), no study has evaluated the effects of different mask coverage levels on face identification accuracy. On the basis of field observations, this study adopted three commonly used mask-wearing methods (each representing different exposed areas of the face) and examined their effects on identification accuracy and time. The results verified the hypothesis by demonstrating that less coverage of mask (e.g., non-FC levels) increased face identification accuracy. However, identification accuracy did not decrease from the MB level to the BB level, indicating that face identification accuracy cannot be further improved by putting the mask down to the BB level. However, identification accuracy can be significantly enhanced by putting the mask down from the FC level to the MB level. This implies that the nose may be one of the crucial clues for face identification, even exposing the upper part of nose can also increase the face identification performance.

Although face identification accuracies in this study were higher than those in previous studies, a large error variance in face matching can

TABLE 3 Tukey's test results of identification accuracy and time.

Sources	Accuracy (%)	Tukey test	Time (s)	Tukey test
Full coverage (FC)	90.3 (8.4)	A	2.82 (0.75)	A
Middle of nose bridge (MB)	93.7 (6.7)	B	2.63 (0.70)	A
Bottom of nose bridge (BB)	94.9 (5.0)	B	2.63 (0.78)	A

Data (mean with standard deviation in parentheses) with the same letter do not differ in Tukey's test.

TABLE 4 One-way ANOVA results of the effects of coverage levels on identification accuracy and time for each sex.

Sex	Response	SS	df	MS	F	value of <i>p</i>	Power
Men	Accuracy	0.081	2	0.040	6.573	<0.01	0.906
	Time	0.935	2	0.468	0.754	0.472	0.177
Women	Accuracy	0.055	2	0.028	7.992	<0.001	0.953
	Time	1.797	2	0.898	1.911	0.151	0.151

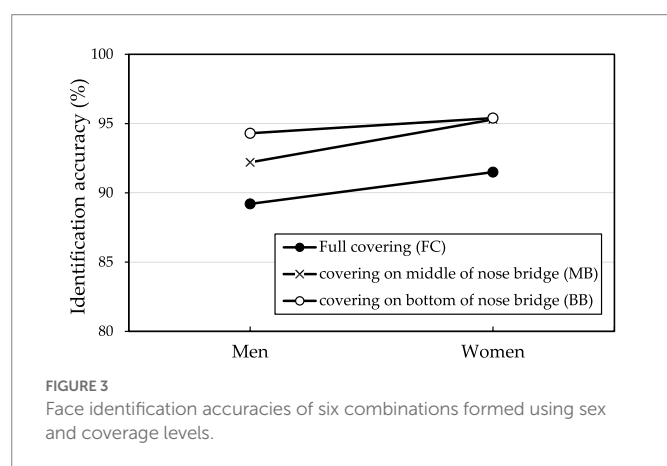


FIGURE 3 Face identification accuracies of six combinations formed using sex and coverage levels.

be explained by the face stimulus characteristics, indicating that specific stimulus features might enhance the identification of masked faces (Estudillo and Bindemann, 2014; Fysh and Bindemann, 2017). Experiments of face identification involved the use of different materials, recognizer characteristics [e.g., super recognizers, Noyes et al., 2021], and testing protocols; therefore, direct comparisons among studies are challenging. The commonly used identification methods in past research included matching (Jenkins et al., 2011; Carragher and Hancock, 2020), recalling (Davies et al., 1977; Burton et al., 2010), image comparison (Megreya and Burton, 2006; Burton et al., 2010), the corresponding target face identified from options (Duchaine and Nakayama, 2006), and shelters (Nguyen and Pezdek, 2017; Noyes et al., 2021; Estudillo and Wong, 2022). Behavioral and neuropsychological studies have also demonstrated that cognitive processes among different approaches reflect distinct cognitive mechanisms (Bindemann and Burton, 2021). Megreya and Burton (2007) also revealed some dissociations between identity match and mismatch trials.

Investigations have determined that matching unfamiliar faces is more challenging than matching familiar ones (Megreya and Burton, 2006). Although this study employed unfamiliar faces as testing subjects, identification accuracy rates were all higher than nearly 90%. This may be because identification of the upper face (Fisher and Cox, 1975; Davies et al., 1977; Dal Martello and Maloney, 2006), specifically the eyes (McKelvie, 1976; Roberts and Bruce, 1988), is more accurate than that of the lower face (e.g., the nose, mouth, and chin). Studies have determined that the eyes are the most crucial cue for face identification. Fisher and

Cox (1975) and McKelvie (1976) observed that the identification of faces with covered eyes was less accurate than that of faces with a covered mouth. Roberts and Bruce (1988) also suggested that covered eyes caused a more significant impact on face identification. Recent studies suggested that covering the face inhibited the recognition of identity and emotional expressions. However, it may also make the eyes more prominent, since they are a reliable index to orient people's social and spatial attention (Villani et al., 2022). In addition, a questionnaire with a forced 4-choice-1-question format was used in this study, and the participants tended to choose the closest answer. Although the participants were not necessarily certain, they recognized that the correct answer existed among the four options, which may also improve identification accuracy.

Identification accuracy and time differed between men and women. A study on female facial expression identification demonstrated that women have faster and more accurate reaction times than men (Lewin and Herlitz, 2002), and also confirmed by other studies (Megreya et al., 2011; Godard et al., 2013; Hansen et al., 2021; Wong and Estudillo, 2022). Lewin and Herlitz (2002) indicated that women's higher face recognition performance was hypothesized to be related to either their higher verbal ability or to their superiority in recognizing female faces. Another face processing study reported that women were better at face recognition than men because they may make more fixations in short fixation durations than men (Herlitz and Rehnman, 2008). Women were faster and more accurate in female facial expression recognition than men, and women looked more at the eyes than men. That is, the superior performance of women in facial expression recognition is related to greater female attention to the eyes, whereas men focus more on the mouth (Hall et al., 2010). The occlusion of the mouth by masks may be a reason for the lower identification accuracy and longer identification time in men than in women. Therefore, the trend of identification accuracy in the BB level differed between men and women (Table 5; Figure 3). This implies that different facial exposure cues caused varying accuracy rates between men and women, therefore, the eye-tracking technique may be a useful tool to clarify the study result in future investigation.

This study has several limitations. Although 115 participants were recruited, the power value of the ANOVA result (sex effect on identification time, Table 2) was less than 0.8, implying the insufficiency of the sample size. This may affect the generalization of the study results. In addition, the overall identification accuracy rates were higher than those in previous studies, which may be because of the forced 4-choice-1-question questionnaire format. An additional option of "no

TABLE 5 Tukey's test results of the effects of coverage levels on identification accuracy and time for each sex.

Men (N=60)		Women (N=55)		
Coverage	Accuracy (%)	Time (s)	Accuracy (%)	Time (s)
FC	89.2 (8.7) A	2.88 (0.73) A	91.5 (6.9) A	2.75 (0.76) A
MB	92.2 (8.2) AB	2.71 (0.74) A	95.3 (5.2) B	2.53 (0.64) A
BB	94.3 (5.3) B	2.75 (0.88) A	95.5 (5.3) B	2.53 (0.65) A

Data (mean with standard deviation in parentheses) with the same letter do not differ in the Tukey's test; FC, full coverage; MB, coverage up to the middle nose bridge; BB, coverage up to the bottom nose bridge.

corresponding face" among the candidate answers may have reduced the possibility of the correct option surmised by the participants, thereby bringing the identification accuracy rate closer to the fact. Furthermore, to control for other interfering variables, we removed the features other than the face used in the test, such as hair, ears, and face shape, resulting in an image that may be slightly different from the actual one. Once other facial features were also appeared, the face may become easier to be identified and needs further verification.

Conclusion

The emergence of the COVID-19 pandemic has led to changes in human communications and interactions. Face masks are commonly used in public spaces. This study examined the effects of three mask coverage levels (FC, MB, and BB) on face identification accuracy and time. The results indicated that identification accuracy and time were significantly affected by sex, and identification accuracy was also influenced by the mask coverage level. During face identification, identification accuracy cannot be further improved by putting the mask at the BB level. Moreover, we recommend wearing masks at the MB level during face identification.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of Chang Gung Memorial Hospital,

Taiwan. Participants provided written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This work was supported by the National Science and Technology Council (NSTC), Taiwan (#110-2221-E-131-025-MY3).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Al Jazeera News (2020). Which countries have made wearing face masks compulsory? Available at: <https://www.aljazeera.com/news/2020/04/countries-wearing-face-masks-compulsory-200423094510867.html>
- Bindemann, M., Attard, J., Leach, A., and Johnston, R. A. (2013). The effect of image pixelation on unfamiliar-face matching. *Appl. Cogn. Psychol.* 27, 707–717. doi: 10.1002/acp.2970
- Bindemann, M., and Burton, M. (2021). "Steps towards a cognitive theory of unfamiliar face matching," in *Forensic Face Matching: Research and Practice*. ed. M. Bindemann (Oxford, England: Oxford University Press), 38–61.
- Bobak, A. K., Mileva, V. R., and Hancock, P. J. (2019). A grey area: how does image hue affect unfamiliar face matching? *Cogn. Res. Princ. Implic.* 4:27. doi: 10.1186/s41235-019-0174-3
- Burton, A. M., White, D., and McNeill, A. (2010). The Glasgow face matching test. *Behav. Res. Methods* 42, 286–291. doi: 10.3758/BRM.42.1.286
- Carragher, D. J., and Hancock, P. J. (2020). Surgical face masks impair human face matching performance for familiar and unfamiliar faces. *Cogn. Res. Princ. Implic.* 5:59. doi: 10.1186/s41235-020-00258-x
- Cartaud, A., Ott, L., Iachini, T., Honoré, J., and Coello, Y. (2020). The influence of facial expression at perceptual threshold on electrodermal activity and social comfort distance. *Psychophysiology* 57:e13600. doi: 10.1111/psyp.13600
- CDC (2020). *Recommendation Regarding the Use of Cloth Face Coverings*, Centers for Disease Control and Prevention, Atlanta, GA.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edn. Erlbaum, Hillsdale, NJ.
- Dal Martello, M. F., and Maloney, L. T. (2006). Where are kin recognition signals in the human face? *J. Vis.* 6, 1356–1366. doi: 10.1167/6.12.2

- Davies, G., Ellis, H., and Shepherd, J. (1977). Cue saliency in faces as assessed by the 'Photofit' technique. *Perception* 6, 263–269. doi: 10.1068/p060263
- Duchaine, B., and Nakayama, K. (2006). The Cambridge face memory test: results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia* 44, 576–585. doi: 10.1016/j.neuropsychologia.2005.07.001
- Estudillo, A. J., and Bindemann, M. (2014). Generalization across view in face memory and face matching. *Iperception* 5, 589–601. doi: 10.1068/i0669
- Estudillo, A. J., Hills, P., and Wong, H. K. (2021). The effect of face masks on forensic face matching: an individual differences study. *J. Appl. Res. Mem. Cogn.* 10, 554–563. doi: 10.1037/h0101864
- Estudillo, A. J., and Wong, H. K. (2022). Two face masks are better than one: congruency effects in face matching. *Cogn. Res. Princ. Implic.* 7:49. doi: 10.1186/s41235-022-00402-9
- Fisher, G., and Cox, R. (1975). Recognizing human faces. *Appl. Ergon.* 6, 104–109. doi: 10.1016/0003-6870(75)90303-8
- Fysh, M. C., and Bindemann, M. (2017). "Forensic face matching: a review" in *Face Processing: Systems, Disorders and Cultural Differences*. eds. M. Bindemann and A. M. Megreya (Hauppauge, NY: Nova Science Publishers), 1–20.
- Ganczak, M., Pasek, O., Duda-Duma, Ł., Świątara, D., and Korzeń, M. (2021). Use of masks in public places in Poland during SARS-CoV-2 epidemic: a covert observational study. *BMC Public Health* 21:393. doi: 10.1186/s12889-021-10418-3
- Godard, O., Baudouin, J. Y., Bonnet, P., and Fiori, N. (2013). Identity–expression interaction in face perception: sex, visual field, and psychophysical factors. *Laterality* 18, 594–611. doi: 10.1080/1357650X.2012.734312
- Graham, D. L., and Ritchie, K. L. (2019). Making a spectacle of yourself: the effect of glasses and sunglasses on face perception. *Perception* 48, 461–470. doi: 10.1177/0301006619844680
- Hall, J. K., Hutton, S. B., and Morgan, M. J. (2010). Sex differences in scanning faces: does attention to the eyes explain female superiority in facial expression recognition? *Cogn. Emot.* 24, 629–637. doi: 10.1080/02699930902906882
- Hansen, T., Zaichkowsky, J., and de Jong, A. (2021). Are women always better able to recognize faces? The unveiling role of exposure time. *PLoS One* 16:e0257741. doi: 10.1371/journal.pone.0257741
- Herlitz, A., and Rehnman, J. (2008). Sex differences in episodic memory. *Curr. Dir. Psychol. Sci.* 17, 52–56. doi: 10.1111/j.1467-8721.2008.00547.x
- Hill, H., and Bruce, V. (1996). The effects of lighting on the perception of facial surfaces. *J. Exp. Psychol. Hum. Percept. Perform.* 22, 986–1004. doi: 10.1037/0096-1523.22.4.986
- Jenkins, R., White, D., Van Montfort, X., and Burton, M. (2011). Variability in photos of the same face. *Cognition* 121, 313–323. doi: 10.1016/j.cognition.2011.08.001
- Kramer, R. S., Young, A. W., and Burton, A. M. (2018). Understanding face familiarity. *Cognition* 172, 46–58. doi: 10.1016/j.cognition.2017.12.005
- Kumar, S., and Lee, H. P. (2020). The perspective of fluid flow behavior of respiratory droplets and aerosols through the facemasks in context of SARS-CoV-2. *Phys. Fluids* 32:111301. doi: 10.1063/5.0029767
- Lewin, C., and Herlitz, A. (2002). Sex differences in face recognition—Women's faces make the difference. *Brain Cogn.* 50, 121–128. doi: 10.1016/S0278-2626(02)00016-7
- McKelvie, S. J. (1976). The role of eyes and mouth in the memory of a face. *Am. J. Psychol.* 89, 311–323. doi: 10.2307/1421414
- Megreya, A. M., Bindemann, M., and Havard, C. (2011). Sex differences in unfamiliar face identification: evidence from matching tasks. *Acta Psychol.* 137, 83–89. doi: 10.1016/j.actpsy.2011.03.003
- Megreya, A. M., and Burton, A. M. (2006). Unfamiliar faces are not faces: evidence from a matching task. *Memory Cogn.* 34, 865–876. doi: 10.3758/BF03193433
- Megreya, A. M., and Burton, M. A. (2007). Hits and false positives in face matching: a familiarity-based dissociation. *Percept. Psychophys.* 69, 1175–1184. doi: 10.3758/BF03193954
- Megreya, A. M., Sandford, A., and Burton, A. M. (2013). Matching face images taken on the same day or months apart: the limitations of photo ID. *Appl. Cogn. Psychol.* 27, 700–706. doi: 10.1002/acp.2965
- Nguyen, T. B., and Pezdek, K. (2017). Memory for disguised same-and cross-race faces: the eyes have it. *Visual Cogn.* 25, 762–769. doi: 10.1080/13506285.2017.1329762
- Niu, Z., Zhou, M., Wang, L., Gao, X., and Hua, G. (2016). Ordinal regression with multiple output CNN for age estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV. 4920–4928.
- Noyes, E., Davis, J. P., Petrov, N., Gray, K. L., and Ritchie, K. L. (2021). The effect of face masks and sunglasses on identity and expression recognition with super-recognizers and typical observers. *R. Soc. Open Sci.* 8:201169. doi: 10.1098/rsos.201169
- Noyes, E., and Jenkins, R. (2017). Camera-to-subject distance affects face configuration and perceived identity. *Cognition* 165, 97–104. doi: 10.1016/j.cognition.2017.05.012
- Rahne, T., Fröhlich, L., Plontke, S., and Wagner, L. (2021). Influence of surgical and N95 face masks on speech perception and listening effort in noise. *PLoS One* 16:e0253874. doi: 10.1371/journal.pone.0253874
- Roberts, T., and Bruce, V. (1988). Feature saliency in judging the sex and familiarity of faces. *Perception* 17, 475–481. doi: 10.1068/p170475
- Siahaan, A. M. P., Lubis, M. P., Dalimunthe, D. A., Nasution, M. R., and Lubis, H. P. R. (2021). Adherence to face mask and social distancing among residents in Medan during the COVID-19 pandemics. *Bali Med. J.* 10, 529–533. doi: 10.15562/bmj.v10i2.2414
- Southall, A., and Van Syckle, K. (2020). Coronavirus bandits? 2 armed men in surgical masks rob racetrack. *The New York Times*. Available at: <https://www.nytimes.com/2020/03/08/nyregion/aqueduct-racetrack-robbery.html>
- Sun, T., Li, L., Xu, Y., Zheng, L., Zhang, W., Zhou, F. A., et al. (2017). Electrophysiological evidence for women superiority on unfamiliar face processing. *Neurosci. Res.* 115, 44–53. doi: 10.1016/j.neures.2016.10.002
- Taiwan Centers for Disease Control [TCDC] (2022). Acts and Regulations. Available at: <https://www.cdc.gov.tw/En/Category/Page/yZOu-4cGeu77HDyzE0ojqg#>
- Villani, C., D'Ascenzo, S., Scerrati, E., Ricciardelli, P., Nicoletti, R., and Lugli, L. (2022). Wearing the face mask affects our social attention over space. *Front. Psychol.* 13:923558. doi: 10.3389/fpsyg.2022.923558
- White, D., Kemp, R. I., Jenkins, R., and Burton, A. M. (2014). Feedback training for facial image comparison. *Psychon. Bull. Rev.* 21, 100–106. doi: 10.3758/s13423-013-0475-3
- Wong, H. K., and Estudillo, A. J. (2022). Face masks affect emotion categorisation, age estimation, recognition, and gender classification from faces. *Cogn. Res.: Princ. Implic.* 7:91. doi: 10.1186/s41235-022-00438-x
- World Health Organization [WHO] (2022). World Health Statistics. Available at: <https://www.who.int/data>
- Zhang, T. T., Zhang, T., and Liu, S. (2022). A modified surgical face mask to improve protection and wearing comfort. *Buildings* 12:663. doi: 10.3390/buildings12050663



OPEN ACCESS

EDITED BY

Alyssa A. Brewer,
University of California,
Irvine, United States

REVIEWED BY

Silvia Spadacenta,
Hertie Institute for Clinical Brain Research,
Germany
Suncica Zdravkovic,
University of Novi Sad,
Serbia

*CORRESPONDENCE

Galina V. Paramei
✉ parameg@hope.ac.uk

SPECIALTY SECTION

This article was submitted to
Perception Science,
a section of the journal
Frontiers in Psychology

RECEIVED 30 May 2022

ACCEPTED 16 January 2023

PUBLISHED 24 February 2023

CITATION

Bimler DL and Paramei GV (2023) Gauging
response time distributions to examine the
effect of facial expression inversion.
Front. Psychol. 14:957160.
doi: 10.3389/fpsyg.2023.957160

COPYRIGHT

© 2023 Bimler and Paramei. This is an open-
access article distributed under the terms of
the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in
other forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Gauging response time distributions to examine the effect of facial expression inversion

David L. Bimler¹ and Galina V. Paramei^{2*}

¹Independent Researcher, Wellington, New Zealand, ²Department of Psychology, Liverpool Hope University, Liverpool, United Kingdom

Introduction: We used images of facial expressions (FEs) of emotion in a speeded *Same/Different* task to examine (i) distributional characteristics of response times (RTs) in relation to inter-stimulus similarity and (ii) the impact of inversion on FE processing.

Methods: Stimuli were seven emotion prototypes, posed by one male and one female, and eight intermediate morphs. Image pairs ($N=225$) were presented for 500ms, upright or inverted, in a block design, each 100 times.

Results: For both upright and inverted FEs, RTs were a non-monotonic function: median values were longest for stimulus pairs of intermediate similarity, decreasing for both more-dissimilar and more-similar pairs. RTs of “Same” and “Different” judgments followed ex-Gaussian distributions. The non-monotonicity is interpreted within a dual-process decision model framework as reflecting the infrequency of identical pairs, shifting the balance between the *Same* and *Different* processes. The effect of stimulus inversion was gauged by comparing RT-based multidimensional scaling solutions for the two presentation modes. Solutions for upright and inverted FEs showed little difference, with both displaying some evidence of categorical perception. The same features appeared in hierarchical clustering solutions.

Discussion: This outcome replicates and reinforces the solutions derived from accuracy of “Different” responses reported in our earlier companion paper. We attribute this lack of inversion effect to the brief exposure time, allowing low-level visual processing to dominate *Same/Different* decisions while elevating early featural analysis, which is insensitive to face orientation but enables initial positive/negative valence categorization of FEs.

KEYWORDS

facial expressions of emotion, inversion effect, *Same/Different* task, response times, dual-process model, ex-Gaussian distribution, multidimensional scaling, cluster analysis

Introduction

Using images of facial expressions (FEs) of emotion, we measured response times (RTs) while participants made “Same”–“Different” judgments on inter-pair similarity. FE pairs were presented both upright and inverted, in the hope that the thus-obtained RT measure would be a sensitive probe of the inversion effect. In the present report, submitted for the Research Topic “Methods and Applications in Perception Science,” we focus on RT distributions as functions of inter-stimulus similarity and the stimulus presentation condition. The aim is to analyze latency of stimulus discriminability, in its relation to an accuracy measure, to further explore processes of perceptual decisions underlying comparisons of visually complex stimuli.

The problem of determining whether two visual stimuli are identical is a natural activity with ecological implications. In experimental psychology, this function is operationalized as the forced-choice *Same/Different* (S/D) task, which has been widely used as a convenient psychometric

technique for measuring (dis)similarities among a number of stimuli (for reviews, see [Farell, 1985, 2022](#)). The *S/D* task has been applied in a range of domains including schematic facial expressions ([Takane and Sergent, 1983](#)); line segments ([Young, 1970](#)); abstract symbols ([Sergent and Takane, 1987](#)); letters ([Podgorny and Garner, 1979](#)); irregular polygons ([Cooper, 1976](#); [Cooper and Podgorny, 1976](#); [Smith et al., 2008](#)); or single-syllable words ([Farell, 2022](#)).

In a *S/D* task, for N stimuli, the $(N^2 - N)$ pairs of different stimuli are each presented some number of times in random order, interspersed with repetitions of the N identical stimulus pairs. The latter provide no similarity information, but in their absence the observers could simply respond “Different” at every trial (though see [Becker, 2012](#); Experiment 2). Non-identical pairs are recognized as such in the majority of trials if exposure times are long enough that inter-stimulus dissimilarities are above the threshold of discrimination. Indices of subjective dissimilarity are the average latency or response time (RT) required to decide that two stimuli differ and the proportion of correct “Different” responses to a given pair (i.e., accuracy).

Assuming that median RTs are a function of subjective dissimilarity, a preliminary to later analysis is to determine the nature of that function. Precedents for this postulate include several studies where RTs were related to inter-stimulus dissimilarities provided directly by subjects in the form of ratings (e.g., [Young, 1970](#); [Podgorny and Garner, 1979](#); [Paramei and Cavanaugh, 1999](#)).

In one widely-accepted form, this postulate states that for trials where different stimuli are correctly recognized as such, the median RT declines steadily as their dissimilarity increases, “an inverse monotonic function between the reaction time data and underlying distances” ([Takane and Sergent, 1983](#), p. 396). This function slopes down steeply when the dissimilarity is subtle, i.e., a small increment in dissimilarity brings a large reduction in the difficulty of decisions, leveling out and approaching a floor value where the difference between the stimuli is immediately apparent ([Cohen and Nosofsky, 2000](#)). Following [Shepard \(1987\)](#), an exponential decline to a constant often fits the function well (e.g., [Paramei and Cavanaugh, 1999](#)).

[Cooper \(1976\)](#) reported that an exponential function fitted RTs from the majority of observers, although a minority appeared to apply a different decision process, and their RTs followed a flat function, not varying significantly with inter-stimulus dissimilarity (see also [Cooper and Podgorny, 1976](#)). Conversely, “just the opposite relation for [incorrect] “same” judgments was experimentally demonstrated [... implying] that for the “same” judgments, reaction time works as a measure of dissimilarity” ([Takane and Sergent, 1983](#), p. 396). A preliminary objective here is to examine the universal truth of the assumption.

Another perspective looks at the entire *distribution* of RTs and responses for a given inter-stimulus dissimilarity, not just the measure of central tendency, and sets out to derive these from first principles ([Balota and Yap, 2011](#)). Of note are random-walk models ([Laming, 1968](#)), the diffusion model ([Ratcliff, 1978, 1985](#)), the race model (e.g., [Huber and O'Reilly, 2003](#)), and others, all falling under the rubric of dual-process decision models. These all agree in postulating two competing “evidence accumulators,” one receptive to any points of difference between the stimuli, and the other to the points on which they agree. These “accumulators” function in parallel until one or other function reaches a threshold. When the evidence for a “Same” decision outweighs that for a “Different” decision, or vice versa, then, depending on the metaphor of choice, the scales tip or the race is won. We will use “*Different*” and “*Same*” to label the decision processes, and “*S*” and “*D*” for the ultimate *response*.

Below, we examine the distributions of RTs for compatibility with dual-process decision models. Naturally this requires a large enough number of trials per stimulus pair (T). Larger values of T also make the average RT more robust, reducing noise from the many hard-to-control variables. There is a trade-off with observer motivation, however, not to mention the danger that observers will learn to recognize each pair as a single Gestalt and provide stereotyped, “over-learned” responses. In previous explorations of RTs as a function of similarity, T has ranged from four (e.g., [Roberson et al., 1999](#), section 5.2), through 10 ([Paramei and Cavanaugh, 1999](#)), up to about 40 or 60 ([Mollon and Cavanaugh, 1986](#)). Much larger values are possible in studies attempting to model the underlying decision mechanisms, which typically examine fewer stimulus pairs. The present study used $T = 100$.

Another aspect of a *S/D* design is the proportion of identical stimulus pairs. The norm for studies in the *S/D* paradigm is to present equal numbers of identical- and different-stimulus trials (e.g., [Qiu et al., 2017](#)). [Smith et al. \(2008\)](#) reported that in a situation with about 50% of identical-pair trials, most human subjects followed a “zero-tolerance” decision strategy, responding *D* to any detectable disparity. In contrast, macaque monkeys appeared to impose a non-zero threshold, responding *S* or *D* to disparities below or above this threshold (as if comfortable with a high number of false-“Same” errors). This can be understood as the *Different* and *Same* processes of a decision model having separate thresholds to attain.

The thresholds can be manipulated by the experimental design: [Ratcliff and Hacker \(1981\)](#) influenced the RTs and relative numbers of *D* and *S* decisions by instructing observers to exercise greater caution before one response or the other. [Downing \(1970\)](#) influenced RTs by manipulating the proportion of trials where the stimuli were identical (50% vs. 25%). In [Smith et al.'s \(2008\)](#) experiment with humans, different-stimulus trials slightly predominated (54%) over same-stimulus pairs. As a precedent, in [Wise and Cain's \(2000\)](#) study about 20% of pairs were identical. Following [Krueger and Shapiro's \(1981, p. 576\)](#) reasoning, it is likely that variation in the ratio of same- and different-stimulus pairs (“heterogeneity of difference”) would shift the perceptual decision criterion. Specifically, they argued that decreasing the proportion of same-stimulus pairs reduces the amount of sensory evidence required to assign the *S* response, thus, resulting in greater number of false-“Same” errors. Indeed, when [Smith et al. \(2008\)](#) manipulated the ratio of identical- vs. different-stimulus pairs (30:35 vs. 40:35) in macaque monkeys, this induced a large shift in tolerance of stimulus disparity in a monkey presented with the 40:35 proportion of identical pairs, i.e., a looser, more inclusive criterion for responding “Same.” This finding is relevant to the present study due to the relatively low proportion of stimulus pairs (7%) which were identical.

In research on FEs of emotion, *S/D* accuracy data have been interpreted as dissimilarities and used to locate the category boundary between distinct emotions ([Calder et al., 1996](#); [Roberson et al., 1999](#); [Suzuki et al., 2005](#)). Such data are also suitable for multivariate analyses such as multidimensional scaling (MDS) and hierarchical cluster analysis, employed to reconstruct the perceptual framework underlying the stimuli, in order to glean clues as to (dynamics of) their cognitive representation.

The present study examined RTs for “Same” and “Different” judgments among images of FEs presented as pairs in upright and in inverted mode. Stimuli contained prototypical posed expressions of emotions and morphed intermediates. We estimated and scrutinized RT functions for individual subjects, while probing the effect of stimulus inversion upon the encoding and processing of FEs in terms of proximities among them in a spatial model. In particular, we asked

whether the inversion impacts more upon some emotions than others; and whether, after inversion, emotion categories still modulate the perception of FEs. A previous MDS analysis of response accuracy in the same experiment (Bimler et al., 2013) found unexpectedly little effect from inversion, and one question we examine here is whether the RTs, as a complementary behavioral measure of (dis)similarity, reveal more effect when examined with the present approach.

Materials and methods

Participants

Two male and two female undergraduate Psychology students, aged 21–25 years, were reimbursed for participation. All participants were right-handed and reported normal vision. Participant sex and poser gender were counterbalanced to offset any possible own-gender bias effect in face recognition (*cf.* Wright and Sladden, 2003). That is, stimuli from the MO series (from a female poser) were presented to one female participant (DK) and one male (HK). Likewise, the WF series (from a male poser) was presented to one female participant (SB) and one male (BF). Each participant completed 30 1-h-long sessions spread over 4 months; for HK and SB these were interrupted by a three-month gap during their summer vacation. The study was conducted in accordance with the ethical principles of the Declaration of Helsinki.

Stimuli

Fourteen grayscale photographs of emotional expressions were selected from *Pictures of facial affect* (Ekman and Friesen, 1976). Those authors deemed these 14 images to be good examples of seven universal emotion categories in unalloyed form [*Happiness* (*H*), *Surprise* (*Su*), *Anger* (*A*), *Sadness* (*Sa*), *Fear* (*F*), *Disgust* (*D*), *Neutral* (*N*)], as evinced by high accuracy of labeling. Seven images featured a female poser identified as MO while the other featured a male poser WF (see Figure 1 in Bimler et al., 2013).

The MO series and WF series were both extended by using image-interpolation software (Design Studio) to create eight ambiguous intermediate stimuli, each lying midway along the continuum defined by two emotion exemplars as end-points. The morphing process involves ‘landmarks’ located within each prototype, allowing smooth interpolation of intermediate stages along a transformation between them (*cf.* Calder et al., 1996; Young et al., 1997). Using morphs between all 21 ($7 \times 6/2$) pairs of “parent” exemplars would have made data collection impractical, so eight pairs were chosen, following a distorted circumplex that paired each exemplar with its neighbor (e.g., *SaN*, *DF*), except for *Anger*, which is paired with *Happiness* (*AH*) and *Surprise* (*ASu*). Each set, MO and WF, included 15 images.

These digitalized stimuli were presented on a 19” CRT-Monitor (V7 N110s), where each image occupied 12.8 cm \times 8.7 cm (subtending $10^\circ \times 6.7^\circ$ at a viewing distance of 74 cm). Measured with a LMT L1009 Luminance Meter, image luminance ranged from 0.23 to 82 cd/m². Ambient lighting in the test room was in the mesopic range (around 10 cd/m²).

Procedure

Each trial consisted of the simultaneous parafoveal presentation of two FE stimuli, symmetrically side-by-side on the screen with a 3.8 cm

gap between them (subtending 3°). After 500 ms, the screen went blank until the participant responded “Same” (*S*) or “Different” (*D*) via a two-button keyboard. Instructions described the stimuli as “emotional faces,” to focus the participants’ attention on their emotional content. Participants were instructed to respond as quickly and correctly as possible. RT was measured (to the nearest 20 ms) from the appearance of the FE pair to the response, by a MS-DOS program running on a Windows-98 PC which controlled presentation and recorded each *S* or *D* response. Each response was followed by an inter-stimulus interval of 300–400 ms, while a small red fixation cross was displayed on the monitor.

In a single run, all possible $15 \times 15 = 225$ pairings of FEs were presented in randomized order. Blocked presentation was used, alternating between blocks of all Upright (*U*) or all Inverted (*I*) pairs. For each participant the experiment began with a practice session of one block in each of the *U* and *I* modes. There followed 10 sessions containing six blocks and 20 containing seven blocks, totalling to 100 runs with FE pairs in the *U* mode and 100 runs in the *I* mode.

Note that 15 of these 225 pairs were indeed identical (7%). The remaining pairs consisted of “left–right” and “right–left” presentations of 105 pairings. These were treated as repetitions in the analysis, ignoring any asymmetry effects. Participants received no indication about how frequently to expect identical pairs, and no feedback about accuracy after trials.

Results

Learning effect

Before further analysis, the data require some rescaling to compensate for any learning effect. Median RTs in each run are plotted for the four participants separately in Figure 1. Clearly these values change across the course of data collection, with some participants showing substantially shorter RTs with accumulated practice (for HK and SB, the abrupt increase in RTs after the 40th trial reflects their summer interruption).

A natural concern is the possibility that RTs for a given pair varied systematically in the course of the experiment relative to other pairs, as subjects became familiar with the stimuli. To test this, we plotted *cumulative sums* of RT_{ij}^* for representative stimulus pairs, across a range of similarities (see Figure 2, exemplified by subject DK, Upright pairs). RT_{ij}^* is defined in the next paragraph. The lines are reasonably straight and do not cross, implying that RTs remained quite stable relative to the median at each run. Thus, Figure 2 shows that changes from learning were across-the-board and did not change the *relationships* among pairs: if a pair evoked a relatively rapid response in the initial runs, it was still relatively rapid at the end of the experiment. However, this progressive change increased the variance of the distribution of RTs for a given pair of FEs.

RTs as a measure of FE inter-stimulus similarity

Following Ratcliff et al. (2010), median RTs for each FE pair were obtained separately for *D* and *S* responses, M_{ij}^D and M_{ij}^S . *U* and *I* mode trials were analyzed separately. The *median* value is preferable to the mean, being unaffected by the skewed distribution of RTs or the outlying, exceptionally delayed responses that sometimes occur. As

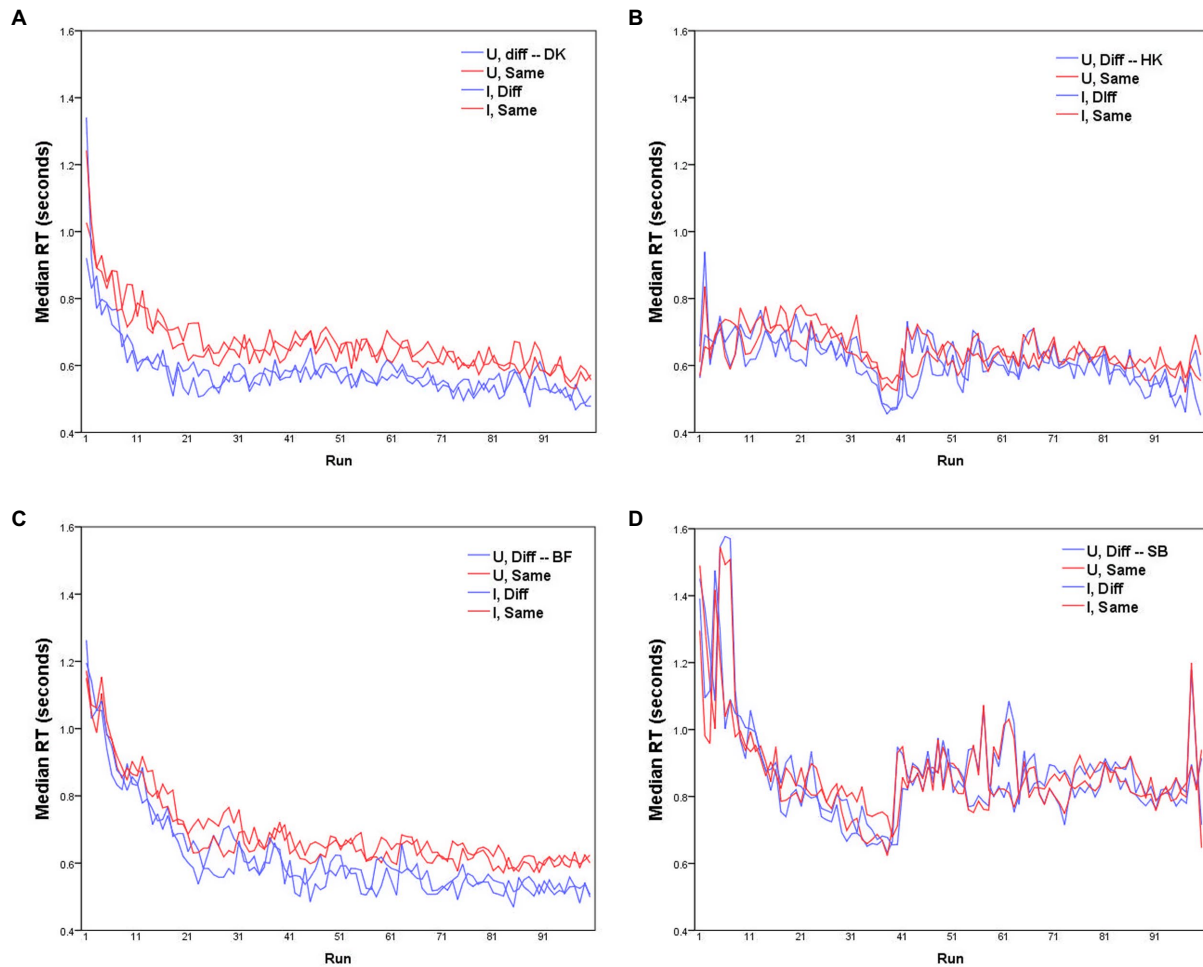


FIGURE 1

Median RTs for each of the four participants (A–D) as a function of run, for *D* responses (blue lines) and *S* responses (red lines), with Upright and Inverted stimuli (solid and dashed lines respectively). MO series: participants (A) DK, (B) HK; WF series: participants (C) BF, (D) SB.

noted, participants' response speed varied as the runs progressed, typically improving with practice. To remove this source of variance, before calculating their medians the 225 RTs in each run r ($1 \leq r \leq 100$) were rescaled with a factor $s(r)$ to bring their median value into line with the global median over all runs for that participant:

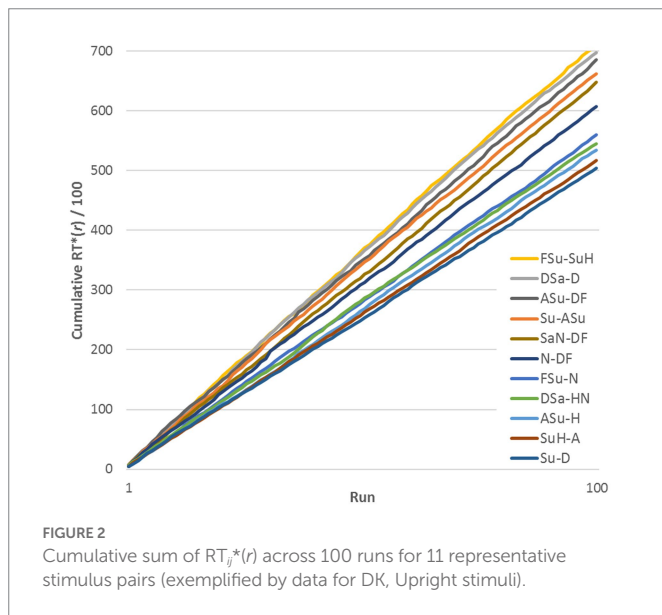
$$RT_{ij}(r) = s(r)RT_{ij}(r)$$

where $s(r) = \text{global median}(RT_{ij}) / \text{median}(RT_{ij}(r))$. We also performed the same analyses without this adjustment, but found no impact on the overall tenor of the outcomes.

Errors occurred relatively often with observers responding *S* in about 25% of the trials (*cf.* 7% of actually identical pairs). The percentage of erroneous *D* judgments for a given pair served as a proxy for the perceived dissimilarity between those stimuli, and was constant across runs. In a companion paper we processed accuracy rate (percentages) of *S* judgments as an index of pairwise perceptual similarity with non-metric multidimensional scaling (MDS) for error-smoothing purposes, embedding them within a four-dimensional geometrical space (Bimler et al., 2013).

RTs to identical FE pairs

First, we explored RTs to identical pairs of FEs, separately for Upright and Inverted conditions. The (objectively) identical pairs of the MO series stimuli most rapidly identified as “Same” were *H-H*, *FSu-FSu*, and *Su-Su*, while the slowest pairs were *N-N*, *San-San*, *ASu-ASu* and *Sa-Sa*. For the WF series, the pairs with the shortest “Same” RTs were *H-H*, *F-F*, and *A-A*, while the slowest pairs were *N-N* and *DF-DF*. These outcomes concur with Becker's (2012) report that in a *S/D* task, matched pairs of negative images took longer to recognize as identical than neutral or positive-affect pairs, a finding attributed to greater demands of processing negative expressions. However, that the most rapidly processed stimuli are marked not so much by their positive affect, but rather by the clarity of a single feature (e.g., WF's exaggerated smile for *Happiness*, or MO's elevated eyebrows and open mouth for *Surprise*). That is, the results are consistent with the observers noticing that the stimuli of a pair share a specific exaggerated feature, apparently tipping the scales toward a *S* response and curtailing further thought. This finding is also in accord with Calvo and Nummenmaa's (2011) conclusion that early (and later) expression discrimination decisions are based on visual saliency of distinctive facial features.



RTs of “Different” vs. “Same” responses to FE pairs

Median RTs for *S* responses (M_{ij}^S) were slightly longer than for *D* responses (M_{ij}^D), as is evident in Figure 3 that plots M_{ij}^D on the horizontal axis against M_{ij}^S for the same pairs on the vertical axis. The slower *S* responses indicate a more conservative decision criterion, in accord with previous findings: *S* responses (conjunctive judgments) imply accumulating more evidence before making the decision whereas for *D* responses (disjunctive judgments) a decision is made as soon as any difference is detected (see Farell, 1985, for a review). The delay in the *S* responses— $\text{mean}(M_{ij}^S - M_{ij}^D)$ —is not constant for all stimulus pairs but varies as a function of pairwise dissimilarity. In addition, $\text{mean}(M_{ij}^S - M_{ij}^D)$ varies from subject to subject, with the largest average delay for BF and DK (90 and 95 ms, respectively) and least for SB (27 ms). It is possible that these inter-individual differences in the delay of *S* responses are spurious, since in RTs of HK (Figure 1B) and SB (Figure 1D) there was an abrupt increase after the summer interruption (we are indebted to a reviewer for this caveat).

RT distributions: Dual-process model

Considering RTs from the dual-process perspective, their distributions become relevant to the possible effects of inversion. In the dual-process paradigm, one can imagine the response to a given stimulus pair (*i, j*) as a Bernoulli model, where the visual system reports to the *Same* and *Different* processes at a regular rate (i.e., at regular clock ticks), and each report has a constant chance of being *Same* and *Different*, this chance depending on how many features the stimuli share (cf. Ratcliff, 1978). The decision process continues until the *Same* or *Different* detector has accumulated enough reports to trigger either a *S* or *D* response, respectively.

Figure 4, using observer DK as example, shows the combined distribution of $RT^*(i, i)$ for FE pairs varying in subjective (dis) similarity, separately for the Upright and Inverted mode. (Results for the other observers are shown in Supplementary Figure S1.) Figure 4A

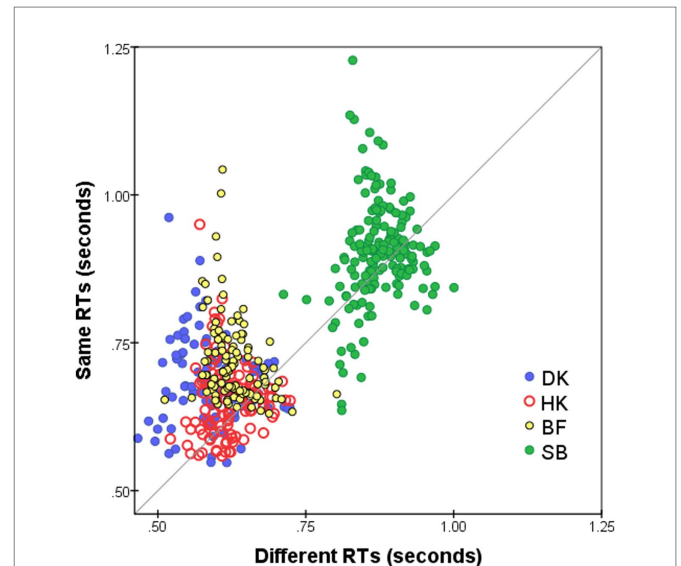


FIGURE 3
Median $RT^* M_{ij}^S$ for *S* responses to *i*-th and *j*-th FEs, $i \neq j$ (vertical axis), plotted against M_{ij}^D for *D* responses (horizontal axis). Superimposed results for four observers and for both Upright and Inverted presentation modes. Pairs omitted if four or fewer responses.

shows the distribution of $RT^*(i, i)$ for the 15 identical-stimulus pairs. These pairs (*i, i*) evoke the *Same* process without interruption from *D* responses and the *S* threshold is almost always reached. Figure 4B shows the combined $RT^*(i, j)$ distributions for 10 pairs which were most distant in DK's accuracy-based MDS solution. Conversely, the *Different* process manifests in isolation in this case of pairs of greatest dissimilarity, where the *D* threshold is almost always reached (RTs for the 15 identical pairs were pooled here, as were the 10 most-dissimilar pairs, to reduce statistical noise in the histograms.) The Bernoulli model predicts *S* and *D* RTs to follow negative binomial distributions, positively skewed, if the accumulation of reports is uninterrupted. As predicted, for DK (and also for the other observers; see Supplementary Figure S1), both distributions are positively skewed with a long “tail” of delayed RTs.

The situation is more complicated for pairs of intermediate dissimilarity. Figures 4C, D plot the $RT^*(i, j)$ distributions for erroneous *S* and correct *D* responses, combining 10 pairs (*i, j*) lying within a band of intermediate distances, chosen so that errors were closest to 50% of responses, i.e., these were pairs for which the dual-process competition was seemingly strongest.

When the reports from the visual system have equal probability of being *S* and *D*, the *Same* and *Different* functions accumulate at only half the rate as in the extreme cases, predicting longer-delayed and therefore less skewed negative binomial distributions. But an additional factor is at play. The probability of a *S* conclusion after some time *t* is reduced by the cumulative probability that a *D* response had already emerged at any time $< t$ (so the theoretical *S* distribution is modulated by the cumulative distribution for the *Different* process). Conversely, the distribution of *D* responses for these pairs is shaped by the cumulative distribution of the *Same* process. Details of this two-way interaction depend on the relative speed of the two processes, among other factors, which might make these specific distributions most sensitive to any effects of inversion.

Past RT data have been successfully modeled by an exponential-Gaussian function with parameters μ , σ , τ (e.g., Heathcote et al., 1991,

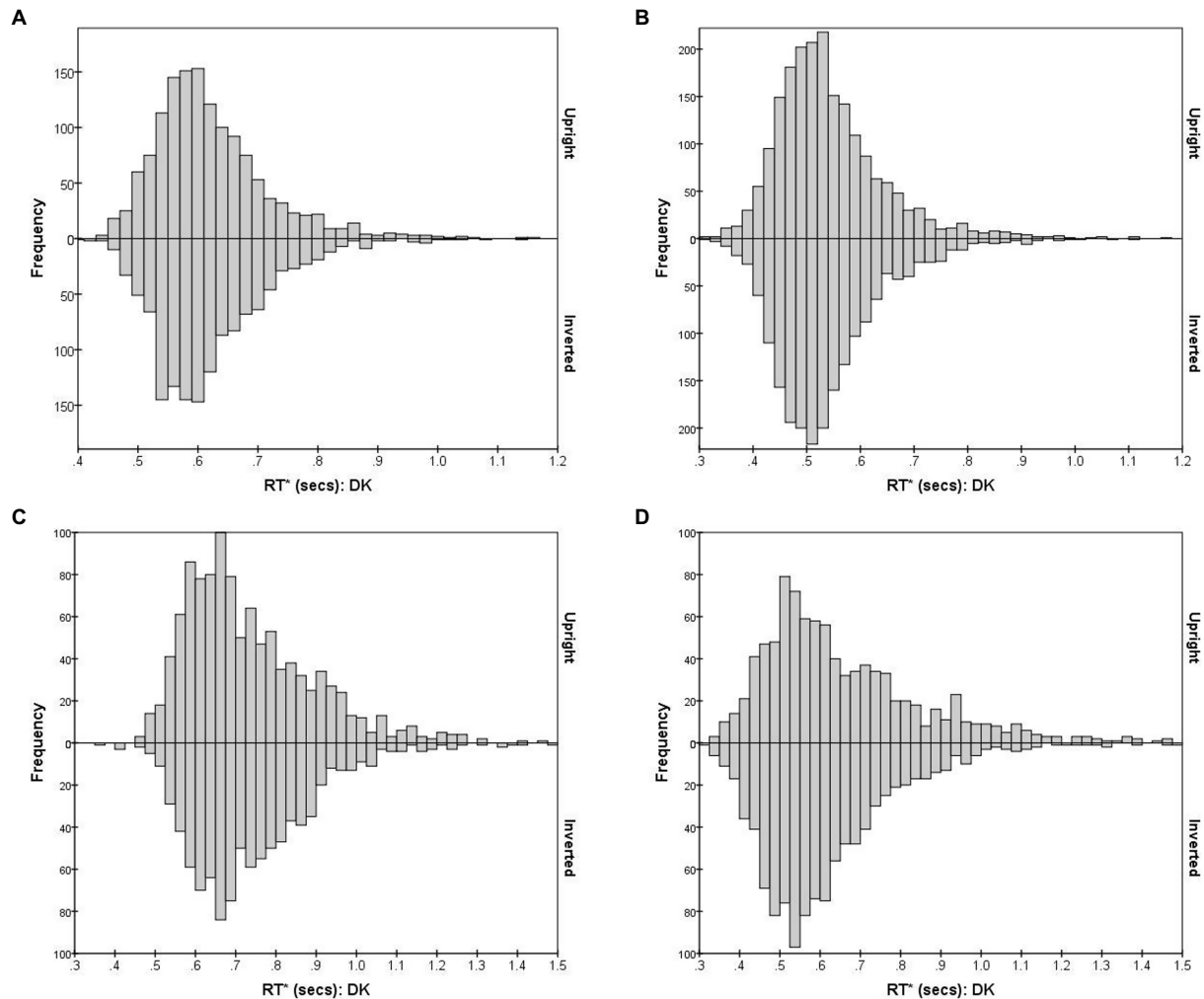


FIGURE 4

RT*(i,j) distributions for observer DK, for FE pairs presented Upright (positive values on the y-axis) and Inverted (negative values on the y-axis). The four graphs illustrate RT*(i,j) distributions for pairs that vary in the degree of inter-stimulus similarity. (A) S responses for 15 identical pairs; (B) D responses for 10 most-different pairs; (C) S responses for 10 intermediate-similarity pairs; (D) D responses for 10 intermediate-similarity pairs.

2019; Balota and Spieler, 1999). Accordingly, we applied the “timefit” function from the “retimes” (package for R), to S response RT*(i,j) values to identical pairs (exemplified by data for participant DK, U mode), as in Figure 4A. The matches between the resulting ex-Gaussian functions and actual distributions for this participant are gratifyingly close (Figure 5), validating the data transformation and suggesting that median values are valid measures of central tendency. Table 1 shows the function parameters and the corresponding moments (mean, standard deviation, skewness) for all observers. (Density functions for the other three observers are presented in Supplementary Figure S2.) Note that in line with previous findings (e.g., Heathcote et al., 1991; Balota and Spieler, 1999), the ex-Gaussian characteristics of individual participants are relatively stable regardless of the mode of FE presentation.

Non-monotonicity of the RT function

Figure 6 plots each observer’s median RTs M_{ij}^D and M_{ij}^S against inter-stimulus dissimilarity (i.e., inter-point distance in that subject’s accuracy-based MDS solution; cf. Bimler et al., 2013). To indicate the

reliability of data-points, the size of each symbol represents the number of decisions on which that median is based.

The unexpected feature is that the M_{ij}^D values follow a peaked function rather than an exponential decline or any other monotonic function (see also Paramei et al., 2009). This is particularly clear in the results for HK, Figure 6B, who took longest to make a D response for an intermediate dissimilarity of about $dist = 6$ (arbitrary units in the MDS solution). As expected, more-distant pairs were judged in less time, but so were more-similar pairs.

The other three observers exhibit comparable non-monotonicity for M_{ij}^D . The M_{ij}^S distributions also follow a non-monotonic contour, though with slightly larger values, the relative delay varying from subject to subject. The shape is less clear because there were very few S responses for highly-dissimilar pairs, so that S points at the right of each panel of Figure 6 are based on only a few atypical responses, limiting their reliability.

We note that SB consistently took 300 ms longer to respond than the other participants (Figure 6D), although the pattern of her RTs is no different. Until further observers are tested, we do not know whether SB is anomalous or at the conservative end of a range of processing-criteria variation, causing a more exhaustive,

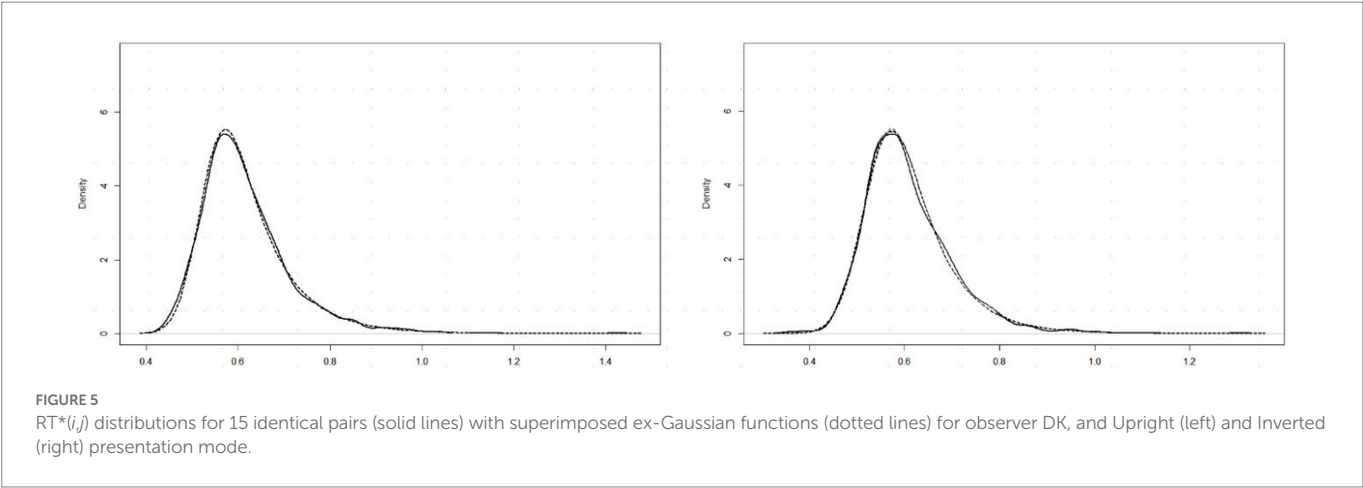


TABLE 1 Parameters (μ , σ , τ) for fitted ex-Gaussian functions and moments (mean, SD, and skewness) for the distributions of S RTs for identical FE pairs in Upright and Inverted mode of presentation, for each observer (Ob).

Mode	Upright						Inverted					
Obs	μ	σ	τ	Mean	SD	Skew	μ	σ	τ	Mean	SD	Skew
DK	0.528	0.043	0.086	0.614	0.096	1.439	0.529	0.048	0.078	0.607	0.092	1.250
HK	0.519	0.046	0.080	0.598	0.092	1.291	0.522	0.043	0.090	0.611	0.099	1.461
BF	0.560	0.044	0.061	0.621	0.075	1.070	0.565	0.046	0.054	0.619	0.070	0.886
SB	0.736	0.067	0.058	0.794	0.089	0.558	0.735	0.056	0.067	0.802	0.087	0.910

attention-strengthened comparison strategy with greater cognitive control (cf. Moret-Tatay et al., 2016).

Filtering RT data prior to calculate MDS solutions

For further analysis we applied MDS to estimates of similarity derived from median RTs. The solution represents stimuli as points in the spatial model, where the proximity of any two points mirrors the corresponding stimulus similarity, and dimensions indicate attributes underlying the perceptual judgments.

To compensate for the non-monotonicity of M_{ij}^D as functions of reconstructed distances, we filtered their values to the range where they were monotonic, by abandoning all entries for stimulus pairs (i,j) that were similar enough for fewer than 33% of trials to return D responses (the exact threshold is not crucial). The result is a *similarity* matrix SIM_D for each observer and each presentation mode, where the matrix elements are $sim_{d_{ij}}$:

$$sim_{d_{ij}} = MD_{ij} \text{ if fraction of } D \text{ responses} < 33\% \\ = [missing\ data] \text{ otherwise}$$

The effect is to retain only pairs from the *right-hand* side of each panel of Figure 6, i.e., the ones that contain information about (sufficiently) large dissimilarities which determine the global structure of MDS solutions. To provide complementary evidence about the finer structure among adjacent stimuli, a second matrix DISS_S was included in the same analysis, consisting of M_{ij}^S values treated as *dissimilarities*—but only for those stimulus pairs where the M_{ij}^D value was rejected, with

[*missing data*] entries otherwise (That is, the entries of this second matrices came from the monotonic *left-hand* half of each M_{ij}^S vs. distance function shown in Figure 6).

$$dissim_{s_{ij}} = MS_{ij} \text{ if fraction of } S \text{ responses} > 66\% \\ = [missing\ data] \text{ otherwise}$$

It follows from the filtering rule that if a stimulus pair (i,j) is represented by its M_{ij}^D value in a filtered SIM_D matrix, while another pair (k,l) is omitted there but is represented by its M_{kl}^S value in the corresponding DISS_S matrix, then (k,l) is more similar than (i,j) . We emphasize that the MDS analysis below does *not* use this inference in any way.

RT-derived MDS solutions

Following accuracy-based analysis in Bilmer et al. (2013), we retained four-dimensional MDS solutions for the MO and WF series separately, for each presentation mode, using an implementation of Kruskal's algorithm in its multiple-matrix repeated-measures mode to pool two subjects' SIM_D and two DISS_S matrices for each poser. Notably, and unexpectedly, no systematic inversion-related differences appeared between solutions for the U and I data, so they are superimposed in Figure 7 as two sets of points, using Procrustes analysis (Gower, 1975) to rotate each pair of solutions to the closest congruence. Given this similarity, we pooled the U and I data to obtain consensus MO and WF solutions (not shown). Values of $Stress_1$ for the 2D to 4D solutions were 0.144, 0.105, 0.088 (MO) and 0.178, 0.126, 0.101 (WF). These $Stress_1$ values and interpretability of all four dimensions justify retention of four dimensions in both cases. After rotation, as expected, the first dimension

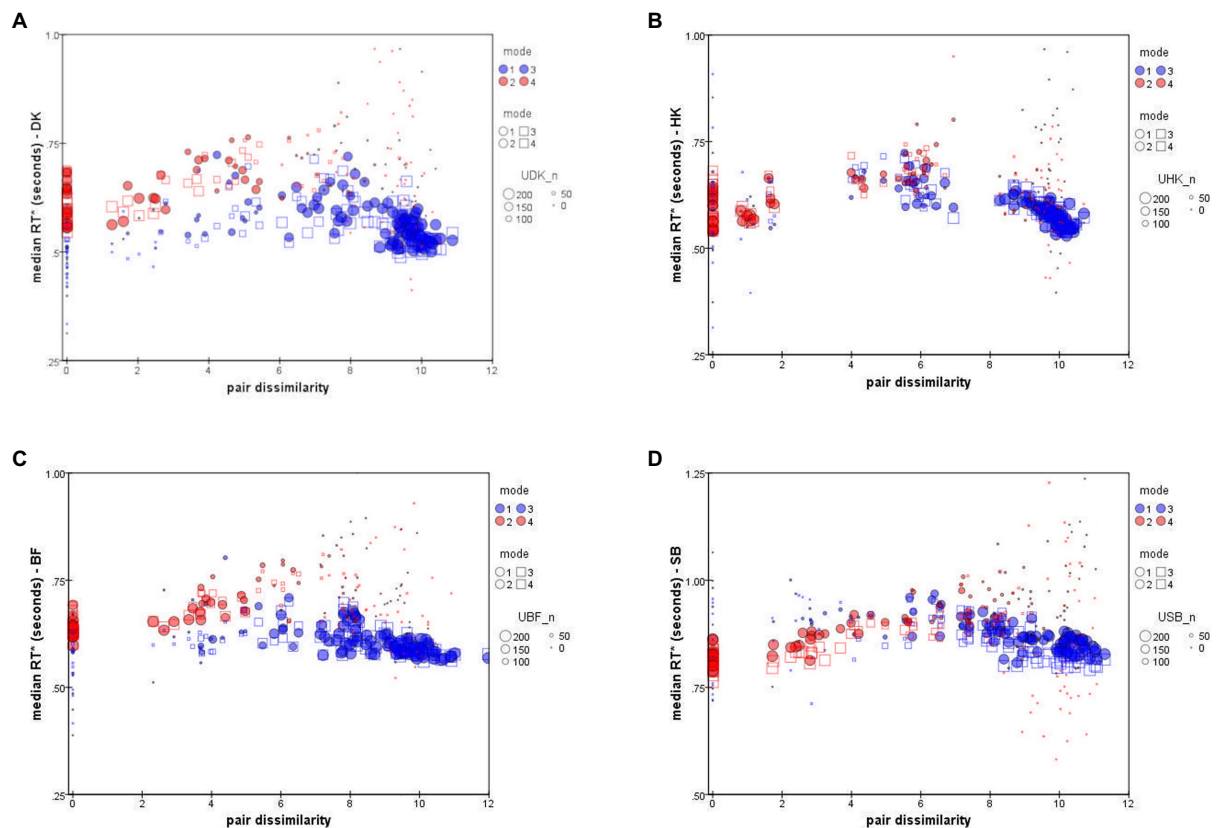


FIGURE 6

Median RT* of “Same” responses, M_{ij}^S (red symbols), and “Different” responses, M_{ij}^D (blue symbols), as a function of distance in the accuracy-rate based MDS solutions. ●=Upright mode; □=Inverted mode. Symbol size represents number of responses to pair (i, j). Data for individual participants: (A) DK, (B) HK, (C) BF, (D) SB (note difference in the y-axis scale).

D1 is a bipolar “Valence” axis, distinguishing the *Happiness* stimulus and its morphs at one extreme from negative-valence FEs at the other. The other axes are unipolar, running from “Neutral” to “Fear/Surprise” (D2), “Anger” (D3) and “Disgust” (D4; see Figure 7).

For confirmation we derived a dissimilarity function (d_{ij}), and created dissimilarity matrices D^{MO} and D^{WF} , by defining the difference between the i -th and j -th stimulus in terms of the profiles of median RTs involving them (the respective rows M_{ik}^D and M_{jk}^D in the similarity matrix). Specifically, d_{ij} is the Euclidean distance between rows M_{ik}^D and M_{jk}^D :

$$d_{ij} = \sqrt{\sum_k (MD_{ik} - MD_{jk})^2}$$

where $i \neq j \neq k$.

Note that in contrast with the original RT data, this dissimilarity function d_{ij} is a monotonic function of the reconstructed distances. Four-dimensional MDS solutions for the MO and WF series (each pooling the matrices for two participants and the two presentation modes) had $Stress_1$ of 0.134 and 0.141, respectively, and were encouragingly similar to those obtained above.

Comparison of solutions derived from RTs and accuracy rates

These combined RT-derived solutions were compared to the solutions extracted from the accuracy-rate data (Bimler et al.,

2013). Similarity between MDS solutions derived from the two behavioral measures was quantified in several ways. One is the Procrustes statistic R^2 , measuring the total sum of residual distances between corresponding points that remain when the configurations have been rescaled, translated, reflected and rotated so as to maximize the overlap between them (Gower, 1975). A value of $R^2 = 0$ indicates complete convergence of the two structures. In this case the values were small: $R^2 = 0.051$ when comparing the solutions from accuracy rates and RTs for the MO series, and $R^2 = 0.032$ for the WF series.

A second form of comparison, canonical correlation (CANCORR), has the advantage of allowing significance tests in the form of Wilks' Λ statistic, here a very stringent test with only 15 points for the correlations. For both the MO and WF series, all four dimensions of the RT solution have recognizable counterparts in the accuracy-rate solution, with $p \leq 0.002$ and $p < 0.005$, respectively.

Effect of FE inversion

No glaring difference between the RTs to Upright and Inverted pairs of stimuli was apparent (Figures 1, 4, 6, 7). At a finer level of analysis, Figure 8(left) plots the D -response median RTs M_{ij}^D for each pair of Inverted stimuli against M_{ij}^D for the identical pair when presented Upright. In the same way Figure 8(right) plots the S -response median RTs M_{ij}^S . It is apparent that inversion failed to substantially affect processing time: the points are concentrated around the diagonal.

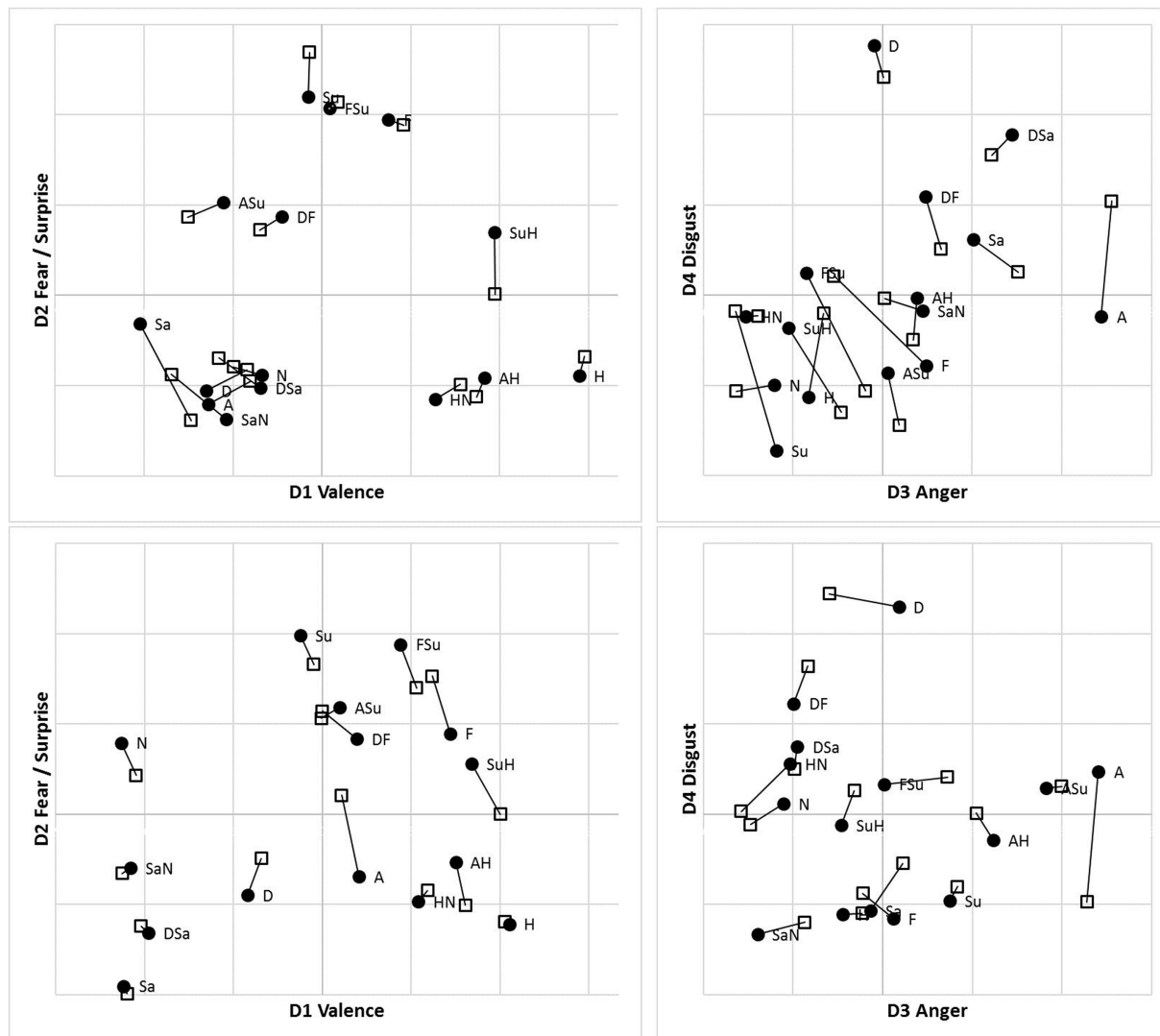


FIGURE 7

Four-dimensional MDS solutions for MO (top) and WF (bottom) stimulus sets, from “filtered” $RT^*(i,j)$ medians sim_d_{ij} and $diss_d_{ij}$, pooling two observers for each stimulus set and superimposing solutions for Upright (●) and Inverted (□) observation (linked by lines). Projection on D1D2 (left) and D3D4 (right) planes.

Significant variations are in the minority (before correcting for multiple comparisons). Median D RTs of participant HK were shorter for Upright pairs than for Inverted pairs by 4.8 ms, $p = 0.029$; in comparison, SB gave faster D and S responses to Inverted than Upright stimuli, by 21.9 and 40.9 ms, respectively, both $p < 0.001$. Crucially, these unsystematic within-participant differences were far smaller than differences between the observers, with DK as the fastest responder and SB as the slowest.

A detailed comparison of 4D structures is difficult when working with 2D perspectives, where a high-dimension rotation can shift points' locations in unexpected ways. Conversely, apparent clustering of points may be coincidental overlaps. To facilitate comparison of the MDS summaries of filtered RTs, we processed the distances within each solution with hierarchical clustering analysis (HCA), specifically, mean link agglomerative algorithm. HCA results (dendrograms) for Upright and Inverted images of the MO and WF series are shown in Figure 9. HCA cannot be applied to the SIM_D and DISS_S matrices directly due to their missing-data entries.

If the subjective similarities represented in MDS solutions strictly follow the construction of the stimulus set, one expects clusters in the

HCA that contain a pair of the prototype FEs and their intermediate morph. For instance, taking *Sadness* and *Neutral* expressions as a relatively similar pair of prototypes, where the *SaN* stimulus is (objectively) between at an equal distance from both, if *SaN* clusters with *Sadness*, then it should be equally similar to *Neutral*, drawing the latter into the cluster. The same might occur for *Fear*, *Surprise* and *FSu*.

For the MO stimuli, as expected, both U and I solutions exhibit a cluster of *Fear/Surprise/FSu*, part of a high-level three-way division (*ASu* is a peripheral member of the cluster, but not the *Anger* prototype). Further, the “negative” emotions *Sadness*, *Anger* and *Disgust* and their morphs comprise a second cluster, joined by *SaN* but not *Neutral*, contrary to the equal-similarity assumption. Finally, the three half-*Happiness* morphs *HN*, *SuH* and *AH* have coalesced with the *Happiness* prototype into the third cluster. That is, the morphs were all more similar to *Happiness* than to their other “parent” prototypes: the response times appear to reflect a degree of categorical processing, which persists when stimuli are inverted. It is worth noting that the same features are present in hierarchical-clustering solutions for accuracy-rate similarity matrices (not shown).

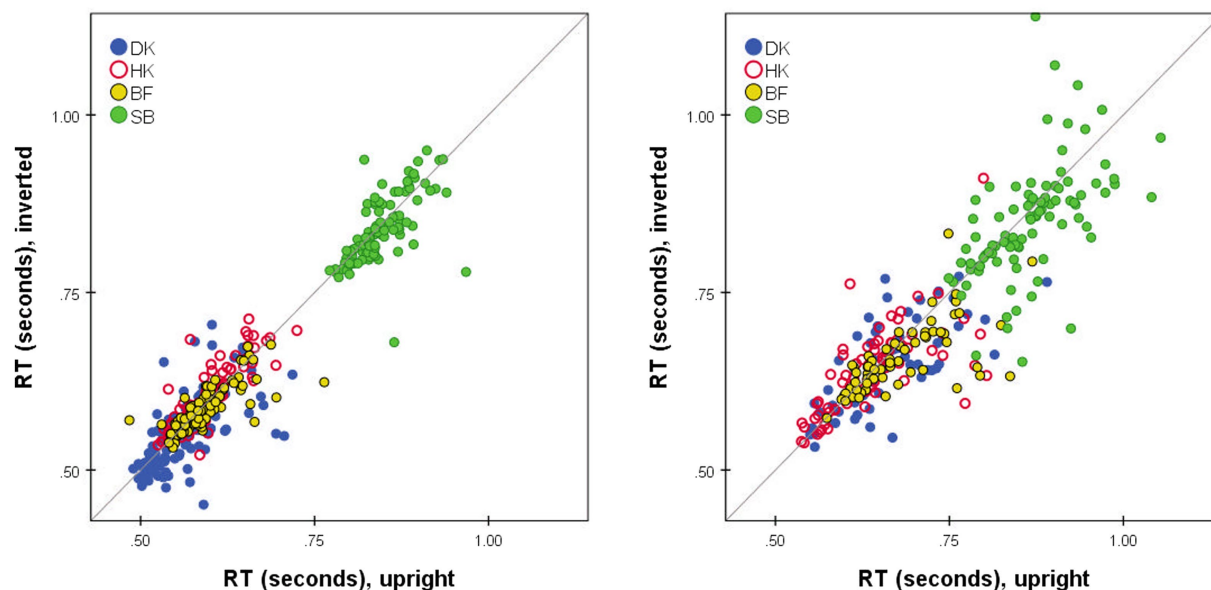


FIGURE 8

Median $RT^*(i,j)$ for Upright pairs (horizontal axis) vs. Inverted pairs (vertical axis). M_{ij}^D for D responses (left); M_{ij}^S for S responses (right). Superimposed results for four observers. Pairs omitted if four or fewer responses in either presentation mode.

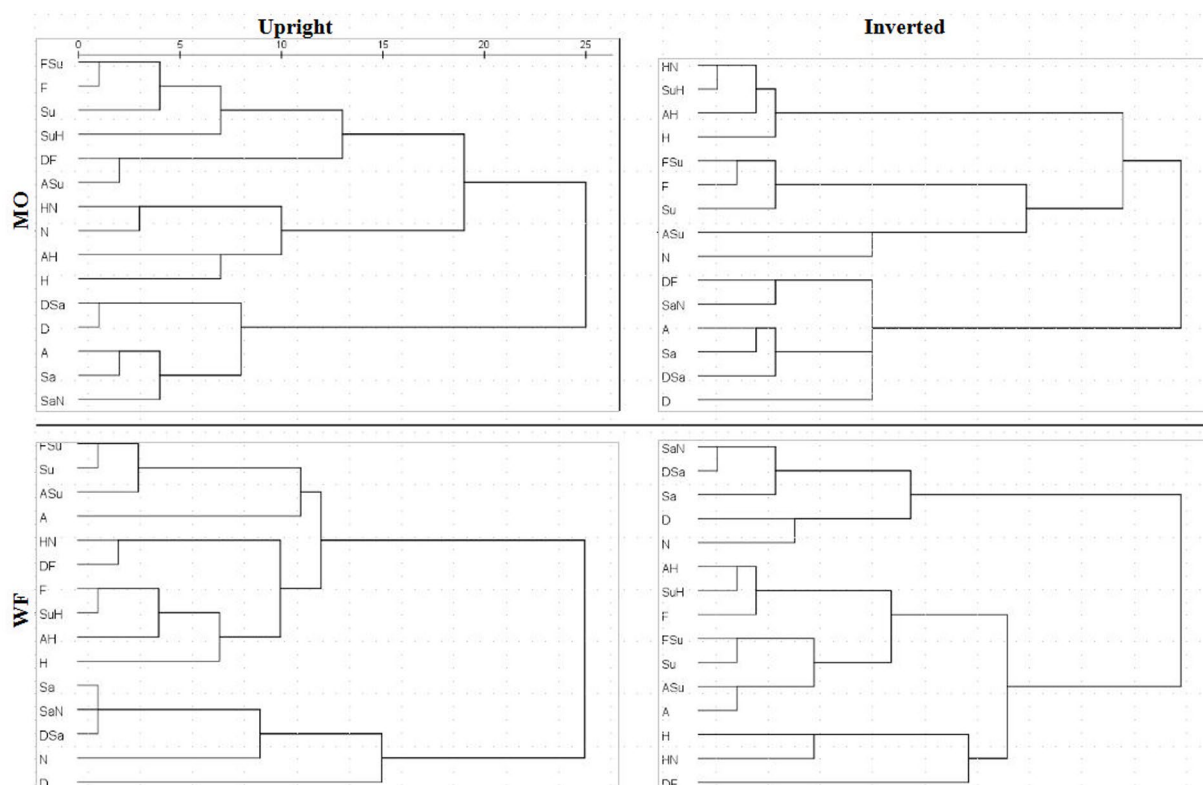


FIGURE 9

Dendrograms derived from MDS solutions for median $RT^*(i,j)$ for MO series (upper row) and WF series (lower row), Upright (left) and Inverted (right).

We observed, however, that inversion *does* shift some morphs from the periphery of one cluster to another. In particular, *SuH* and *DF* are both in the *Fear/Surprise* group according to Upright responses, but

inversion shifts *SuH* to the *Happiness* cluster and *DF* to the “negative” cluster. Conversely, *Neutral* is in the *Happiness* cluster according to Upright responses (linked by its proximity to *HN*), but is in the *Fear/*

Surprise group when inverted, where the connection is less obvious. One cannot read too much into these details, as the cluster membership of ambiguous stimuli is susceptible to random fluctuations in the data.

The WF solutions, in comparison, exhibit a two-way split, distinguishing a cluster of *Sadness*, *Disgust*, *Neutral* and their morphs. Other stimuli are more of a continuum, affected by data fluctuations. The *Happiness* prototype and morphs do not coalesce completely for Inverted data, but *Fear*, *AH* and *SuH* do form a consistent tight cluster. Another tight cluster can be discerned in both U and I results, combining *Surprise* with *FSu* and *ASu*, along with *Anger* (drawn by proximity to *ASu*) but not *Fear*. Recall that these local details are driven by the DISS_S matrices, in contrast to the SIM_D matrices determining the global structure.

Discussion

Learning effect

Following the outline of the Results section, we begin by addressing possible learning effect on RTs. The unusually large number of trials and responses in this study ($T=100$ per stimulus pair) highlight certain important aspects of RTs. First, they vary with practice—generally becoming shorter, albeit with relapses. This will not surprise anyone who has played computer games or learned to touch-type, engraining automatic motor pathways by dint of time and practice. Importantly though, these fluctuations were not accompanied by any systematic shift in accuracy rates across the course of data collection, as assessed by the discrimination d' and bias C parameters (Bimler et al., 2013). It would seem that the improvements are limited to the motor skill of pressing one or the other key. Crucially for present purposes, RTs for specific pairs show no progressive changes over runs when rescaled to be proportional to the median RT for a given run.

This learning effect is not a serious concern here because the drift affects the median RT to the same extent across stimulus pairs. However, extensive practice does increase the *variance* around each median, obscuring the shape of the distributions, which provides the rationale for reducing that variance with the corrections applied here. Presumably similar practice shifts also occurred, although to a lesser extent, in RT data gathered with fewer *S/D* trials (e.g., Takane and Sergent, 1983).

Individual variation

The variations among observers are of note. Even within the small group studied here, SB responded more slowly than the others with a smaller relative advantage for *D* responses (Figures 1D, 3), and a relative advantage for Inverted pairs (Figure 8). It may be that personality traits influence these differences in decision and, hence, dual-process parameters. For example, Nowicki and Cooley (1990) reported that subjects with internal locus of control were significantly faster to distinguish different emotions in a *S/D* task. Cognitive style may, also, underlie differences in an individual's strategy of attention allocation, reflected by an eye-movement pattern of exploration of the compared images that affects the process and chronometry of a perceptual decision (for a discussion cf. Bendall et al., 2016).

The ex-Gaussian functions fitted to the RT data (Figure 5; Table 1) may be relevant here. The μ and σ parameters of the normal

component characterize the “leading edge” of a distribution and are conjectured to reflect early automatic processing, whereas τ , the exponential component characterizing the length of the tail, is more likely to reflect central attentional process (e.g., Balota and Spieler, 1999). As seen in Table 1, μ and σ are higher for SB than for other participants in both U and I modes while τ is lower, i.e., distribution skewness. Moret-Tatay et al. (2016) attribute lower τ to reducing either the tendency to double-check by a participant before responding or the amount of attentional lapsing. One can speculatively interpret SB's lower τ as fewer attentional lapses than other participants, i.e., her “attentional-based strengthening” due to “enhancement in cognitive control” (in the authors' terms) was higher, also increasing μ and σ . However, as noted by Heathcote et al. (1991) and Matzke and Wagenmakers (2009), the rationale for ex-Gaussian functions is more empirical than theoretical, so we have not attempted to interpret the parameters further.

The role of the ratio of same- and different-stimulus pairs in the design

Another factor that possibly affected the obtained outcome is the imbalance of (factually) same- and different-stimulus pairs in our experimental design. The identical-stimulus pairs in the present study, 7%, were scarce compared to the default design in RT research of 50% of total trials. This aspect of the design, facilitated by the brief presentation time, might have encouraged observers to overlook points of difference (or to pay more attention to points of similarity), biasing their judgments toward erroneous *S* responses (cf. Krueger and Shapiro, 1981). It may be that in the absence of explicit guidance for how many “same” pairs to expect, and in the absence of corrective feedback, participants implicitly set their own targets for what seemed a plausible rate and adjusted the thresholds of their decision processes accordingly. In the dual-process framework, the present design is likely to have had an impact on the drift rate of information accrual in the dual processes (cf. Ratcliff, 1978) and, hence, the balance between them. More specifically, it seems to have elevated the critical level of stimulus disparity required for an *S* or *D* judgment to be equally likely (with observers accepting disparities below this criterion as ‘sufficiently identical’), a strategy regarded by Smith et al. (2008) as characteristic of non-human primates. The predominance of different-stimulus pairs appears to have the unintended effect of increasing the number of erroneous *S* responses, with the benefit of enhancing the statistical robustness of false *S* rates when they serve as an index of stimulus similarity. As a result, we were previously able to use each subject's percentage of *D* responses to each stimulus pair as a yardstick of “dissimilarity” (Bimler et al., 2013).

We also observed that *S* judgments were made more slowly than *D* judgments for the same-stimulus pair (Figure 3), although the delay varies with pair and with the participant. This may be another outcome of the change of processing criteria, in accordance with Downing's (1970) report that *S* responses were slower when they were less frequent (25% vs. 50% of trials). Future studies could examine the effects of variation in the proportion of same- and different-stimulus pairs on shifting the criterion of perceptual decisions, potentially reflected by accuracy rate and mean RT difference of *S* and *D* responses.

The RT function of stimulus similarity does not sustain monotonicity

The outcome of the observers' relaxed criterion of "sameness" in the present study revealed a feature of the *Same/Different* task that normally is obscured. In particular, the data become a test of the assumption that RTs are a monotonic function of stimulus dissimilarity, assessed by accuracy ("%D"; Bimler et al., 2013) or by distances in a RT-derived MDS solution as a smoothed version, as here. As demonstrated in the Results section, in either case, despite the simplicity and intuitive appeal of the assumption, it fails for the present data.

We argue, with Ratcliff (1985) and also Ratcliff et al. (2010), that this non-monotonic RT/dissimilarity relationship is in fact an inherent feature of the task, usually concealed but brought to the foreground by aspects of the present study. When the experimental design allows a non-identical stimulus pair to attract a substantial number of *S* responses, it becomes possible for the *Same* process to forestall the *Different* process, and thereby truncate the distribution of RTs for *D* responses. For relatively similar pairs, the *S* responses become the majority, and they are also brief (because their underlying RT function is *decreasing* with similarity). Thus, *D* responses are recorded only on the trials when the *Different* process happens to handle the visual information even more rapidly—probably because some unmistakable, visually salient point of difference between the stimuli "popped out" during their presentation.

Non-monotonicity is a natural corollary of a dual-process decision model. We argue that it is not observed in studies where there are 50% of same-stimulus pairs with zero disparity which attract the majority of *S* responses, these being correct. In such cases the response threshold for the *Same* process arguably is higher than in the present data (or the threshold for the *D* response is lower). This shifts the cross-over between the two processes—the level of dissimilarity where either response to a pair is equally likely, and the *D*-response distribution is truncated—to a difference too small to appear in any non-identical stimulus pairs.

If, as proposed above, the observers here have adjusted their decision criteria and handicapped the *Different* process, as it were—shifting their thresholds to increase the fraction of *S* responses—this would produce Figure 6 as a side-effect. Note that Podgorny and Garner (1979), who reported RTs for *D* responses to same-stimulus pairs, indeed found that, consistent with our prediction, these incorrect *D* responses were actually *shorter* than RTs for many different-stimulus pairs, presumably because of the *Same* process forestalling longer responses. The principle with which we began—that "reducing dissimilarity increases the median value of the distribution of *Different* RTs"—is only true in the special case that the dual processes of the decision model can operate in isolation. This account predicts that RTs for *D* responses will peak at the dissimilarity for which the two kinds of responses are equally common, for that observer. Inspection of the data shows this to be the case.

The same argument further predicts that median RTs of *S* responses will also be a non-monotonic function of dissimilarity, because at larger dissimilarities where the *Different* process operates rapidly we only see the truncated lower tail of the *S* distribution, from those trials where the *Same* process has operated more rapidly still. As noted, this was the case for DK. Such responses are rare, however. There is an additional complication that a *S* response to a dissimilar pair can also occur if the observer was unable to respond promptly (due, e.g., to inattention during the 500 ms of presentation),

leaving neither process with adequate information beyond a fading memory trace, and forcing a delayed and effectively random response. Consistent with this hypothesis, here RTs were generally long when subjects gave *S* responses for dissimilar pairs (often exceeding 1 s). Such "timed-out" trials in this scenario can also yield long *D* responses, but they are lost in the large majority of rapid *D* responses, having little impact on M_{ij}^D .

MDS solution derived from a corrected RT function

When the non-monotonicity is recognized it can be corrected, making the data suitable for MDS. Several lines of evidence converge to validate the RT-derived MDS solutions obtained here. In particular, these solutions concur with geometric models for the same stimuli obtained previously by interpreting the accuracy rate as a dissimilarity measure (Bimler et al., 2013). Furthermore, all present RT-derived solutions are plausible with regard to both formal and explanatory MDS criteria: they have low values of badness-of-fit ($Stress_1$); their dimensions lend themselves to straightforward interpretation as continuous affective gradients; and they are internally consistent as models of the relationships among FE stimuli (specifically, the point representing each morph is located somewhere between the points for the prototype FE "parents"). This validates the use of median RTs as a measure of perceptual difference (in particular, for probing category effects) in the more common situation where the error rate is *not* high enough to serve as an index of proximity. Here we concur with Wise and Cain (2000, p. 261), who summed *S/D* errors across subjects and across bands of stimulus pairs, to support their conclusion that "latency to discriminate shows promise as an objective measure of qualitative similarity."

No effect of inversion on discrimination of facial expressions

An unexpected result was that inversion of stimulus pairs had no consistent effect on response times (Figures 1, 7) or the relative order of RTs (Figures 6, 8), despite our expectation that the emotional content of inverted FEs would require slower serial analysis of local features due to the disruption of processing configural cues, whereby distal features are integrated into a unified whole (Rossion, 2008). One obvious explanation is that, in the challenging discrimination task, the affective content was simply not involved in the *S/D* decisions, these being made purely on the basis of similarity of face trivial details or visually salient diagnostic features (Arnold and Lipp, 2011; Calvo and Nummenmaa, 2011; Murphy et al., 2020; Baldassi et al., 2022), and increase in contrast (in the mouth region for happiness and the eye region for fear; Psalta and Andrews, 2014) or of perceived *physical* similarity, with the stimuli undergoing a low-level form of comparison as abstract patterns of gray-tones and textures (cf. Bimler et al., 2013).

However, there is some evidence of an effect of inversion upon RTs and accuracy rates, but this is confined to pairs of intermediate dissimilarity where the competition between the dual processes is strongest, and thus susceptible to small changes in their parameters. Comparing the Upright and Inverted RT*(*i,j*) mode distributions for intermediate-similarity pairs for DK indicates that this observer identified more of the sufficiently-similar pairs as *D* when Inverted (Figure 4D) and (erroneously) as *S* when upright (Figure 4C). Responses

of HK followed the same pattern, with inversion increasing the median *D* RT (Supplementary Figure S1). However, this subtle difference is not consistent across observers, and for SB the effect of inversion was to increase the number of correct *D* responses (Supplementary Figure S1, third row), in addition to allowing faster responses than in Upright presentation (Figure 8). These pairs were outnumbered by pairs that were sufficiently similar or sufficiently dissimilar for the *Same* or *Different* process to operate without interference, and they were lost in the MDS solutions.

Feature-based early extraction of FE affective meaning

The weakness of any FE inversion effect is in accord with previous findings in two studies measuring response speed—one using a visual search paradigm (Lipp et al., 2009) and another on the choice of a face with the highest emotional content in a horizontally aligned pair (Baldassi et al., 2022), where both suggest a role for low-level processing of face images. Even so, the present study demonstrates partial extraction of affective information despite the brief stimulus presentation. Specifically, the “Valence” dimension is extracted early in visual processing of facial expressions, in line with previous studies using visual search (Lipp et al., 2009) or forced-choice paradigms (for a review see Calvo and Nummenmaa, 2016). As argued by Calvo and Nummenmaa (2011, p. 1758), it is possible that affective information is extracted at some point but is “only minimally used due to [it] being overshadowed by the earlier extracted, and simpler to be managed, visual saliency information, which then was retained also for later discrimination stages.” This conclusion is in accord with findings in an ERP study demonstrating that an emotional expression effect is recorded as early as 120–180 ms post-stimulus (Eimer and Holmes, 2007).

The exposure of 500 ms may have been too short for the processing of the configural information involved in the full decoding of affective meaning (cf. Rossion, 2014; Calvo and Nummenmaa, 2016; Murphy et al., 2020) but it was evidently long enough to extract low-level visual cues at early stages of processing face expressions, shown to be manifested in ERPs within 200 ms (Batty and Taylor, 2003; Ashley et al., 2004; Ohmann et al., 2016; Duan et al., 2022) or relatively correct affective estimates of FEs presented for 150 ms (Arnold and Lipp, 2011). A further corollary can provisionally be drawn from the Results, that some sufficiently low-level facial property distinguished *Sadness*, while no such property was unique to *Fear* or *Surprise* (any low-level cues being shared between those last two expressions).

Early *Happiness* categorical perception effect

Further, along with the subjective dichotomy of the expression “Valence,” our results reveal an early categorical perception (CP) effect for *Happiness*-images (cf. Bilmer and Kirkland, 2001). Although the nature of the stimulus set precludes a rigorous test of CP, the RT-based dissimilarities do display certain hallmarks of CP. In particular, some of the interpolated morphed *Happiness*-stimuli subjectively are not midway between their two constituent prototypes, but are displaced toward one (i.e., are harder to distinguish from it). The *Anger-Happiness*, *Surprise-Happiness* and *Happiness-Neutral* morphs (*AH*, *SuH*, and *HN*) were all closer to the *Happiness* prototype than to their other constituent,

although the *HN* morph (for instance) is physically just as close to *Neutral*. This would be the outcome if they fell within the boundary of the *Happiness* category. Notably, this structure of the subset of *Happiness*-stimuli is present in separate MDS solutions, congruent with our previously-reported accuracy-based outcomes of this study (Bimler et al., 2013). The emerged *Happiness*-subsets in the RT-derived MDS solutions are in accord with outcomes of cluster analysis (see Figure 9 in the present paper). In addition, it is buttressed by Fechner analysis outcomes of our accuracy data: *K*-means clustering revealed a *H*, *AH*, *SuH*, and *HN* cluster, for both U and I conditions and all participants (Dzhafarov and Paramei, 2010). The early emerging *Happiness* category concurs with the finding of the N170 component that reflects earliest manifestation of the CP effect (Qiu et al., 2017; Duan et al., 2022).

The *Happiness* categorization effect is also visible in the hierarchical clustering extracted from Upright and Inverted RTs (Figure 9). The *AH*, *SuH*, and *HN* morphs combine with the *Happiness* prototype in a distinct cluster: in speeded similarity decisions, the 50% of *Happiness* dominates the 50% of other emotional prototypes, regardless of inversion. This early manifestation of the *Happiness* category concurs with both the timing of psychophysiological components—a shorter latency of the N170 (Batty and Taylor, 2003) and categorization advantage (lower discriminability) of positive expressions (Qiu et al., 2017); and, as well, of behavioral measures of responses to *Happy* faces—shorter visual search times (Lipp et al., 2009), saccade latencies (Calvo and Nummenmaa, 2011; Beaudry et al., 2014), and faster choices of the “happiest” than the “angriest” face in a pair (Baldassi et al., 2022).

Our finding is also in accordance with the visual salience of a “smile” feature estimated using behavioral measures: Smith and Schyns (2009) and Bombardi et al. (2013) found the *Happiness* expression to be low-spatial-frequency rich, involving the mouth as its distinguishing feature more than the other prototype FEs. Lower perceptual thresholds for *Happiness* detection compared with *Fear*, *Anger* or *Sadness* point out to the lip-end raise as a highly diagnostic feature (Calvo and Nummenmaa, 2011; Du and Martinez, 2013; Maher et al., 2014; Calvo et al., 2018). This distinctive single cue bypasses integration of face parts (cf. Calvo et al., 2014) and, in the present study, renders the categorical processing of briefly glimpsed facial *Happiness*, in both its “pure” and morphed forms, less susceptible to inversion.

Concluding remarks

The absence of difference in the RT pattern of “Same” – “Different” judgments between Upright and Inverted FE pairs suggests that the participants were viewing the (briefly presented) faces not as Gestalts but, rather, as abstract patterns of low-level features (maybe as gray-tone gradients). In the above discussion of potential information-processing mechanisms behind the present findings we favored the explanation that leans upon the dual-process model and implies separate thresholds in the *Same* and *Different* accruing processes.

This explanation, though, is not the only possibility. We are grateful to an anonymous reviewer, who pointed out that this finding of no inversion effect implies that the *S/D* task is a search paradigm, whereby independent (low-level) stimulus components undergo an analytic and self-terminating comparison, a “feature search” or search of feature conjunctions (Farell, 1985, 2022). Our results (Figure 3) fit into a serial search paradigm with expected shorter *D* responses compared to *S* responses. However, as pointed out by Farell (1985) in his seminal work,

the relation between the pattern of “Same” – “Different” judgments and parallel vs. serial processing is difficult to determine.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Institute of Psychology, Darmstadt University of Technology, Germany. The participants provided their written informed consent to participate in this study.

Author contributions

DB conducted the data analysis and contributed to writing the manuscript. GP contributed to the study design, running the experiment, initial data processing, and writing the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgments

The authors are grateful to Slawomir J. Skwarek and Sina Holtfreter (néé Willems) for indispensable assistance in data collection and Wolfgang Bösch for adapting the software for stimulus generation. Valuable advice on the experimental design from Ehtibar N. Dzhafarov and Hans Colonius is gratefully acknowledged. We thank the

participants for their time, understanding and collaborative spirit. Helpful comments of two reviewers are highly appreciated.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2023.957160/full#supplementary-material>

SUPPLEMENTARY FIGURE S1

RT* distributions, Upright and Inverted, for HK (top row), BF (middle row) and SB (bottom row). (a) S responses for 15 identical pairs; (b) D responses for 10 most-different pairs; (c) S responses for 10 intermediate-similarity pairs; (d) D responses for 10 intermediate-similarity pairs.

SUPPLEMENTARY FIGURE S2

RT*(i,j) distributions for 15 identical pairs (solid lines) with superimposed ex-Gaussian functions (dotted lines) for three remaining participants (rows), in Upright (left) and Inverted (right) modes of FE presentation.

References

- Arnold, D. H., and Lipp, O. V. (2011). Discrepant integration times for upright and inverted faces. *Perception* 40, 989–999. doi: 10.1068/p6955
- Ashley, V., Vuilleumier, P., and Swick, D. (2004). Time course and specificity of event-related potentials to emotional expressions. *Neuroreport* 15, 211–216. doi: 10.1097/00001756-200401190-00041
- Baldassi, G., Murgia, M., Prpic, V., Rigutti, S., Domijan, D., Agostini, T., et al. (2022). Attentional capture in emotion comparison is orientation independent. *Psychol. Res.* doi: 10.1007/s00426-022-01683-x
- Balota, D. A., and Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: beyond measures of central tendency. *J. Exp. Psychol. Gen.* 128, 32–55. doi: 10.1037/0096-3445.128.1.32
- Balota, D. A., and Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry: the power of response time distributional analyses. *Curr. Dir. Psychol. Sci.* 20, 160–166. doi: 10.1177/0963721411408885
- Batty, M., and Taylor, M. J. (2003). Early processing of the six basic facial emotional expressions. *Cogn. Brain Res.* 17, 613–620. doi: 10.1016/s0926-6410(03)00174-5
- Beaudry, O., Roy-Charland, A., Cormier, I., and Tapp, R. (2014). Featural processing in recognition of emotional facial expressions. *Cognit. Emot.* 28, 416–432. doi: 10.1080/02699931.2013.833500
- Becker, M. W. (2012). Negative emotional photographs are identified more slowly than positive photographs. *Atten. Percept. Psychophys.* 74, 1241–1251. doi: 10.3758/s13414-012-0308-x
- Bendall, R. C. A., Galpin, A., Marrow, L. P., and Cassidy, S. (2016). Cognitive style: time to experiment. *Front. Psychol.* 7:1786. doi: 10.3389/fpsyg.2016.01786
- Bimler, D., and Kirkland, J. (2001). Categorical perception of facial expressions of emotion: evidence from multidimensional scaling. *Cognit. Emot.* 15, 633–658. doi: 10.1080/02699930143000077
- Bimler, D., Skwarek, S. J., and Pamei, G. V. (2013). Processing facial expressions of emotion: upright vs. inverted images. *Front. Psychol.* 4:54. doi: 10.3389/fpsyg.2013.00054
- Bombardi, D., Schmid, P. C., Schmid Mast, M., Birri, S., Mast, F. W., and Lobmaier, J. S. (2013). Emotion recognition: the role of featural and configural face information. *Q. J. Exp. Psychol.* 66, 2426–2442. doi: 10.1080/17470218.2013.789065
- Calder, A. J., Young, A., Perrett, D., Etcoff, N., and Rowland, D. (1996). Categorical perception of morphed facial expressions. *Vis. Cogn.* 3, 81–118. doi: 10.1080/713756735
- Calvo, M. G., Fernández-Martín, A., and Nummenmaa, L. (2014). Facial expression recognition in peripheral versus central vision: role of the eyes and the mouth. *Psychol. Res.* 78, 180–195. doi: 10.1007/s00426-013-0492-x
- Calvo, M. G., Gutiérrez-García, A., and Del Libano, M. (2018). What makes a smiling face look happy? Visual saliency, distinctiveness, and affect. *Psychol. Res.* 82, 296–309. doi: 10.1007/s00426-016-0829-3
- Calvo, M. G., and Nummenmaa, L. (2011). Time course of discrimination between emotional facial expressions: the role of visual saliency. *Vision Res.* 51, 1751–1759. doi: 10.1016/j.visres.2011.06.001
- Calvo, M. G., and Nummenmaa, L. (2016). Perceptual and affective mechanisms in facial expression recognition: an integrative review. *Cognit. Emot.* 30, 1081–1106. doi: 10.1080/02699931.2015.1049124
- Cohen, A. L., and Nosofsky, R. M. (2000). An exemplar-retrieval model of speeded same-different judgments. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 1549–1569. doi: 10.1037/0096-1523.26.5.1549
- Cooper, L. A. (1976). Individual differences in visual comparison processes. *Atten. Percept. Psychophys.* 19, 433–444. doi: 10.3758/BF03199404
- Cooper, L. A., and Podgorny, P. (1976). Mental transformations and visual comparison processes: effects of complexity and similarity. *J. Exp. Psychol. Hum. Percept. Perform.* 2, 503–514. doi: 10.1037/0096-1523.2.4.503
- Downing, B. D. (1970). Response probabilities and “same-different” reaction times. *Atten. Percept. Psychophys.* 9, 213–215. doi: 10.3758/BF03212631

- Du, S., and Martinez, A. M. (2013). Wait, are you sad or angry? Large exposure time differences required for the categorization of facial expressions of emotion. *J. Vis.* 13:13. doi: 10.1167/13.4.13
- Duan, Y., Zhan, J., Gross, J., Ince, R. A. A., and Schyns, P. G. (2022). Network mechanisms of dynamic feature selection for flexible visual categorizations. arXiv:2205.04393v2. arXiv [Preprint].
- Dzhafarov, E. N., and Pamei, G. V. (2010). "Space of facial expressions: cumulated versus transformed dissimilarities" in *Fechner day 2010. Proceedings of the 26th annual meeting of the international society for psychophysics*. eds. A. Bastianelli and G. Vidotto (Padua: University of Padova), 605–610.
- Eimer, M., and Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia* 45, 15–31. doi: 10.1016/j.neuropsychologia.2006.04.022
- Ekman, P., and Friesen, W. V. (1976). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists Press.
- Farell, B. (1985). "Same"–"different" judgments: a review of current controversies in perceptual comparisons. *Psychol. Bull.* 98, 419–456. doi: 10.1037//0033-2909.98.3.419
- Farell, B. (2022). Hypothesis testing, attention, and 'Same'-'Different' judgments. *Cogn. Psychol.* 132:101443. doi: 10.1016/j.cogpsych.2021.101443
- Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika* 40, 33–51. doi: 10.1007/BF02291478
- Heathcote, A., Lin, Y. S., Reynolds, A., Strickland, L., Gretton, M., and Matzke, D. (2019). Dynamic models of choice. *Behav. Res. Methods* 51, 961–985. doi: 10.3758/s13428-018-1067-y
- Heathcote, A., Popiel, S. J., and Mewhort, D. J. K. (1991). Analysis of response time distributions: an example using the Stroop task. *Psychol. Bull.* 109, 340–347. doi: 10.1037/0033-2909.109.2.340
- Huber, D. E., and O'Reilly, R. C. (2003). Persistence and accommodation in short-term priming and other perceptual paradigms: temporal segregation through synaptic depression. *Cognit. Sci.* 27, 403–430. doi: 10.1207/s15516709cog2703_4
- Krueger, L. E., and Shapiro, E. G. (1981). A reformulation of Proctor's unified theory for matching-task phenomena. *Psychol. Rev.* 88, 573–581. doi: 10.1037/0033-295X.88.6.573
- Laming, D. R. J. (1968). *Information theory of choice reaction time*. New York, NY: Wiley.
- Lipp, O., Price, S. M., and Tellegen, C. L. (2009). No effect of inversion on attentional and affective processing of facial expressions. *Emotion* 9, 248–259. doi: 10.1037/a0014715
- Maher, S., Ekstrom, T., and Chen, Y. (2014). Greater perceptual sensitivity to happy facial expression. *Perception* 43, 1353–1364. doi: 10.1068/p7806
- Matzke, D., and Wagenmakers, E.-J. (2009). Psychological interpretation of ex-Gaussian and shifted Wald parameters: a diffusion model analysis. *Psychon. Bull. Rev.* 16, 798–817. doi: 10.3758/PBR.16.5.798
- Mollon, J. D., and Cavonius, C. R. (1986). The discriminability of colours on C.R.T displays. *J. I. Electron. Rad. Eng.* 56, 107–110. doi: 10.1049/jiere.1986.0036
- Moret-Tatay, C., Leth-Steensen, C., Irigaray, T. Q., Argimon, I. I. L., Gamermann, D., Abad-Tortosa, I., et al. (2016). The effect of corrective feedback on performance in basic cognitive tasks: an analysis of RT components. *Psychol. Belg.* 56, 370–381. doi: 10.5334/pb.240
- Murphy, J., Gray, K. L. H., and Cook, R. (2020). Inverted faces benefit from whole-face processing. *Cognition* 194:104105. doi: 10.1016/j.cognition.2019.104105
- Nowicki, S., and Cooley, E. L. (1990). The role of locus of control orientation in speed of discriminating facial affect. *J. Res. Pers.* 24, 389–397. doi: 10.1016/0092-6566(90)90029-6
- Ohmann, K., Stahl, J., Mussweiler, T., and Kedia, G. (2016). Immediate relativity: EGG reveals early engagement of comparison in social information processing. *J. Exp. Psychol. Gen.* 145, 1512–1529. doi: 10.1037/xge0000222
- Pamei, G. V., Bilmer, D. L., and Skwarek, S. J. (2009). "Comparing reaction times and accuracy rates in a same/different task with facial expressions" in *Fechner day 2009. Proceedings of the 25th annual meeting of the international society for psychophysics*. eds. M. A. Elliott, S. Antonijevic, S. Berthaud, P. Mulcahy, C. Martyn and B. Bargery et al. (Galway, Ireland: National University of Ireland), 251–254.
- Pamei, G. V., and Cavonius, C. R. (1999). Color spaces of color-normal and color-abnormal observers reconstructed from response times and dissimilarity ratings. *Atten. Percept. Psychophys.* 61, 1662–1674. doi: 10.3758/BF03213125
- Podgorny, P., and Garner, W. R. (1979). Reaction time as a measure of inter- and intraobject visual similarity: letters of the alphabet. *Atten. Percept. Psychophys.* 26, 37–52. doi: 10.3758/BF03199860
- Psaltis, L., and Andrews, T. J. (2014). Inversion improves the recognition of facial expression in thatcherized images. *Perception* 43, 715–730. doi: 10.1068/p7755
- Qiu, R., Wang, H., and Fu, S. (2017). N170 reveals the categorical perception effect of emotional valence. *Front. Psychol.* 8:2056. doi: 10.3389/fpsyg.2017.02056
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev.* 85, 59–108. doi: 10.1037/0033-295X.85.2.59
- Ratcliff, R. (1985). Theoretical interpretations of speed and accuracy of positive and negative responses. *Psychol. Rev.* 92, 212–225. doi: 10.1037/0033-295X.92.2.212
- Ratcliff, R., and Hacker, M. J. (1981). Speed and accuracy of same and different responses in perceptual matching. *Atten. Percept. Psychophys.* 30, 303–307. doi: 10.3758/BF03214286
- Ratcliff, R., Thapar, A., and McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cogn. Psychol.* 60, 127–157. doi: 10.1016/j.cogpsych.2009.09.001
- Roberson, D., Davidoff, J., and Braisby, N. (1999). Similarity and categorisation: neuropsychological evidence for a dissociation in explicit categorisation tasks. *Cognition* 71, 1–42. doi: 10.1016/S0010-0277(99)00013-X
- Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. *Acta Psychol. (Amst)* 128, 274–289. doi: 10.1016/j.actpsy.2008.02.003
- Rossion, B. (2014). Understanding face perception by means of human electrophysiology. *Trends Cogn. Sci.* 18, 310–318. doi: 10.1016/j.tics.2014.02.013
- Sergent, J., and Takane, Y. (1987). Structures in two-choice reaction-time data. *J. Exp. Psychol. Hum. Percept. Perform.* 13, 300–315. doi: 10.1037/0096-1523.13.2.300
- Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science* 237, 1317–1323. doi: 10.1126/science.3629243
- Smith, J. D., Redford, J. S., Haas, S. M., Coutinho, M. V. C., and Couchman, J. J. (2008). The comparative psychology of same-different judgments by humans (*Homo sapiens*) and monkeys (*Macaca mulatta*). *J. Exp. Psychol. Anim. Behav. Process.* 34, 361–374. doi: 10.1037/0097-7403.34.3.361
- Smith, F. W., and Schyns, P. G. (2009). Smile through your fear and sadness: transmitting and identifying facial expression signals over a range of viewing distances. *Psychol. Sci.* 20, 1202–1208. doi: 10.1111/j.1467-9280.2009.02427.x
- Suzuki, A., Shibui, S., and Shigemasa, K. (2005). "Temporal characteristics of categorical perception of emotional facial expressions" in *Proceedings of the twenty-sixth annual conference of the cognitive science society, 4–7 august 2004*. eds. K. Forbus, D. Gentner and T. Regier (Hillsdale, NJ: Erlbaum), 1303–1308.
- Takane, Y., and Sergent, J. (1983). Multidimensional models for reaction times and same-different judgments. *Psychometrika* 48, 393–423. doi: 10.1007/BF02293683
- Wise, P. M., and Cain, W. S. (2000). Latency and accuracy discriminations of odor quality between binary mixtures and their components. *Chem. Senses* 25, 247–265. doi: 10.1093/chemse/25.3.247
- Wright, D. B., and Sladden, B. (2003). An own gender bias and the importance of hair in face recognition. *Acta Psychol. (Amst)* 114, 101–114. doi: 10.1016/s0001-6918(03)00052-0
- Young, F. W. (1970). Nonmetric scaling of line length using latencies, similarity, and same-different judgments. *Atten. Percept. Psychophys.* 8, 363–369. doi: 10.3758/BF03212609
- Young, A. W., Rowland, D., Calder, A. J., Etcoff, N. L., Seth, A., and Perrett, D. I. (1997). Facial expression megamix: tests of dimensional and category accounts of emotion recognition. *Cognition* 63, 271–313. doi: 10.1016/s0010-0277(97)00003-6



OPEN ACCESS

EDITED BY

Anil Ufuk Batmaz,
Concordia University,
Canada

REVIEWED BY

Fernando Marmolejo-Ramos,
University of South Australia,
Australia
Xiaohua Cao,
Zhejiang Normal University,
China

*CORRESPONDENCE

Merle Sagehorn
✉ merle.sagehorn@uni-osnabrueck.de

SPECIALTY SECTION

This article was submitted to
Perception Science,
a section of the journal
Frontiers in Psychology

RECEIVED 22 September 2022

ACCEPTED 27 February 2023

PUBLISHED 28 March 2023

CITATION

Sagehorn M, Johnsdorf M, Kisker J, Sylvester S,
Gruber T and Schöne B (2023) Real-life
relevant face perception is not captured by the
N170 but reflected in later potentials: A
comparison of 2D and virtual reality stimuli.
Front. Psychol. 14:1050892.
doi: 10.3389/fpsyg.2023.1050892

COPYRIGHT

© 2023 Sagehorn, Johnsdorf, Kisker, Sylvester,
Gruber and Schöne. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).
The use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Real-life relevant face perception is not captured by the N170 but reflected in later potentials: A comparison of 2D and virtual reality stimuli

Merle Sagehorn^{1*}, Marike Johnsdorf¹, Joanna Kisker¹,
Sophia Sylvester², Thomas Gruber¹ and Benjamin Schöne¹

¹Experimental Psychology I, Institute of Psychology, Osnabrück University, Osnabrück, Germany,

²Semantic Information Systems Research Group, Institute of Computer Science, Osnabrück University, Osnabrück, Germany

The perception of faces is one of the most specialized visual processes in the human brain and has been investigated by means of the early event-related potential component N170. However, face perception has mostly been studied in the conventional laboratory, i.e., monitor setups, offering rather distal presentation of faces as planar 2D-images. Increasing spatial proximity through Virtual Reality (VR) allows to present 3D, real-life-sized persons at personal distance to participants, thus creating a feeling of social involvement and adding a self-relevant value to the presented faces. The present study compared the perception of persons under conventional laboratory conditions (PC) with realistic conditions in VR. Paralleling standard designs, pictures of unknown persons and standard control images were presented in a PC- and a VR-modality. To investigate how the mechanisms of face perception differ under realistic conditions from those under conventional laboratory conditions, the typical face-specific N170 and subsequent components were analyzed in both modalities. Consistent with previous laboratory research, the N170 lost discriminatory power when translated to realistic conditions, as it only discriminated faces and controls under laboratory conditions. Most interestingly, analysis of the later component [230–420ms] revealed more differentiated face-specific processing in VR, as indicated by distinctive, stimulus-specific topographies. Complemented by source analysis, the results on later latencies show that face-specific neural mechanisms are applied only under realistic conditions (A video abstract is available in the [Supplementary material](https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1050892/full#video-abstract) and via YouTube: <https://youtu.be/TF8wiPUrpsY>).

KEYWORDS

face perception, N170, virtual reality, late potentials, realistic conditions, ecological validity, EEG

1. Introduction

As an inherently social species, humans highly rely on their ability to appraise faces. The human brain is specialized in recognizing and interpreting faces, as cortical regions, like, e.g., the fusiform face area and cells in the inferior temporal cortex, are especially sensitive to face stimuli (Mccarthy et al., 1997; Nestor et al., 2008; Pyles et al., 2013;

Weiner and Zilles, 2016; Palejwala et al., 2021), enabling the extraction of social information as well as the reaction to, and the interaction with the social environment. The cognitive and emotional mechanisms underlying these abilities have evolutionarily evolved in a socially complex and responsive environment. Specifically, they have been attuned to social situations in which at least two individuals are present and engaging in face-to-face communication. A person directly addressing another communicates information and usually expects a verbal or non-verbal response (Keltner and Kring, 1998; Keltner et al., 2003). Thus, the physical proximity preconditioning social involvement facilitates the facial information to immediately become self-relevant. When being actively involved in a social situation, it is imperative to extract affective information from facial expressions to deduce the other persons' emotions, intentions and expectations (Keltner and Kring, 1998; Hamm et al., 2003; Matsumoto et al., 2008). A situation in which a person has to process facial expressions conveying emotions and intentions that do not hold any relevance, virtually do not occur in real-life and thus do not correspond to the normal operating mode of neuronal face recognition mechanisms. In a real-life situation, for example, it is impossible for anger, reflected by facial expressions, to not possess interpersonal meaning for the recipient.

However, this rather observational approach, where participants are confronted with faces that do not bear any or only little meaning, constitutes the conventional laboratory paradigm employing 2D-monitor presentation of faces. To put it differently, in the conventional laboratory participants observe faces, but even though these faces look directly at them, the participants are not seen by anyone. Under real-life conditions, this is a highly improbable scenario. For the sake of experiential control, facial stimuli are even further reduced to their basic physical attributes. Conventional laboratory experiments study face perception by means of gray scale pictures usually eliminating head shape and hair style (e.g., [Jemel et al., 2003](#); [Miyoshi et al., 2004](#); [Blau et al., 2007](#); [Dering et al., 2011](#)), and investigate effects of inversion, i.e., faces presented upside down (e.g., [Rossion et al., 1999](#); [Itier and Taylor, 2004](#); [Vizioli et al., 2010](#)). Paying tribute to the complexity of real-life face perception, more ecological valid approaches use colored pictures of whole scenes ([Rousselet et al., 2004](#)), dynamic faces and face animations ([Recio et al., 2011](#)) or videos featuring faces ([Johnston et al., 2015](#)). Still, typical face stimuli are not only limited in realism concerning properties such as resolution, color, perspective and size. They are also not part of an egocentric reference frame – only the laboratory setup they are being presented in is – and as aforementioned devoid from any social context. Therefore, face perception might only be examined as an isolated process. Given the apparent discrepancies between real-life and laboratory setups, the neural mechanisms observed in a 2D environment might be domain specific and not exhibit the same functional properties in a realistic setting.

Virtual Reality (VR) is technically capable of increasing the realism of classical laboratory designs by allowing for presentation of stimuli in real-world size, offering depth structure and spatial proximity ([Parsons, 2015](#); [Pan and Hamilton, 2018](#); [Snow and](#)

[Culham, 2021](#); [Schöne et al., 2021b](#); [Kisker et al., 2021b](#)). Previous studies employing VR as a method for direct comparison of cognitive and emotional mechanisms under conventional laboratory as opposed to realistic conditions found significant deviations between the two modalities. Long-standing effects established by various classical laboratory experiments could not be replicated in VR, respectively ([Schöne et al., 2021a,b](#); [Kisker et al., 2021c](#)). Specifically, replicating [Simons and Chabris \(1999\)](#) seminal invisible gorilla paradigm in VR revealed that inattention blindness plays a much more subordinate role than the original experiment might have implicated ([Schöne et al., 2021b](#)). Not only attentional processes change their operational mode under realistic conditions, but also memory encoding and retrieval work differently. The well-established theta old/new-effect only occurs when remembering pictorial stimuli, whereas recognition of scenes where the participant was actually present relies on different mnemonic processes ([Kisker et al., 2021b](#)). Also, a comparison of emotional and motivational markers by means of frontal alpha asymmetries provides evidence that the models derived from laboratory data cannot be applied to realistic settings without restrictions ([Schöne et al., 2021a](#); [Kisker et al., 2021c](#)). These differences between laboratory and VR settings can be ascribed to the fact that VR is technically capable to create a highly realistic and thus self-relevant experiences (e.g., [Gabana et al., 2018](#); [Schöne et al., 2019, 2021a](#); [Kisker et al., 2021b](#); [Newman et al., 2022](#)). Perceived realism in virtual environments can be derived from observing behavior during virtual experiences (e.g., [Blascovich et al., 2002](#); [Gromer et al., 2018](#); [Xu et al., 2021](#); [Kisker et al., 2021a,c](#)) as well as neural responses ([Schöne et al., 2023](#)). Moreover, the autobiographic mnemonic mechanisms guiding retrieval as their employment suggest that those experiences are remembered as if they were real ([Schöne et al., 2019, 2021a](#); [Kisker et al., 2021c](#)).

So far, only few studies have investigated face perception using VR, focusing on context modulations ([Stolz et al., 2019](#)) and emotional valence encoding in face perception ([Kirasirova et al., 2021](#)). Up to our knowledge, however, none have included the comparison between laboratory and realistic conditions. Thus, as a first step towards a more holistic understanding of face perception in an ecological valid setting, we translated the classical laboratory setup into a VR setting, enhancing realism with the goal of maintaining strict experimental control ([Parsons, 2015](#); [Snow and Culham, 2021](#)). To this end, we presented images of people on a 2D monitor to the participants in a blocked within-design, and the very same scene as an immersive 3D virtual experience in which the presented people sat directly in front of the participants.

To ensure maximal comparability of the results obtained under VR and 2D conditions and to be able to integrate them into the vast body of electrophysiological scientific literature on face perception, we followed the overall rationality of laboratory conventions, i.e., sequential randomized presentation of stimuli in a controlled environment. Although the sequential presentation of up-popping static persons in the physical vicinity of the participants is physically still highly improbable, this setup bridges the gap between the conventional laboratory and a more realistic approach to face perception tackling the issue of neglecting involvement and self-relevance.

Most importantly, it allows for comparing the canonical event-related potentials (ERPs) associated with face perception. The most discussed neural correlate of perceptual processing of human faces is the ERP component N170, which has been related to face perception and categorization based on observed stimulus-dependent amplitude differences (Rossion and Jacques, 2011). It is characterized by a negative deflection in amplitude occurring at about 170 ms after presentation of a face that can be measured at occipito-temporal electrode positions, i.e., over posterior visual cortical areas (Eimer, 2011; Rossion and Jacques, 2011). Stronger, i.e., more negative, N170 amplitudes have been shown to occur for faces compared to objects and for emotional faces compared to neutral faces (e.g., Itier and Taylor, 2004; Blau et al., 2007). Still, further results showed that the N170 is also sensitive to other objects (e.g., cars; Dering et al., 2009; Boehm et al., 2011), dependent on expertise (e.g., Bukach et al., 2006; Ip et al., 2017), also influenced by perceptual variety of the stimulus material (e.g., Thierry et al., 2007; Dering et al., 2009; Boehm et al., 2011) and not able to differentiate human and ape faces (Zion-Golumbic and Bentin, 2007). Source localization of the N170 points to the fusiform gyrus (McCarthy et al., 1997; Herrmann et al., 2005b; Rossion and Jacques, 2011) and the superior temporal sulcus region (STS; Itier and Taylor, 2004) while brain activation patterns additionally involve sources in a parieto-temporal-occipital network (Herrmann et al., 2005b).

Another component potentially relevant in face perception is the P1, an early positive component occurring at around 100 ms post-stimulus at occipito-temporal electrodes. Albeit the P1 has been shown to mostly be sensitive to various low-level perceptual properties of visual stimuli (e.g., stimulus contrast), direction of spatial attention and arousal state (Rossion et al., 1999; Jemel et al., 2003), it has also been associated with face perception, as some studies report a categorial sensitivity towards faces (Itier and Taylor, 2002; Herrmann et al., 2005a; Thierry et al., 2007; Dering et al., 2009; Kuefner et al., 2010).

Modulation by social relevance (Bublitzky et al., 2014), contextual and self-related emotion (Herbert et al., 2013; Stolz et al., 2019) as well as decision-relevant information (Ratcliff et al., 2009), can only be observed in later components (> 250 ms). While the N170 component is clearly involved in basic face categorization to some degree, more profound processing resulting in a global, individualized and highly informative face representation is manifested in other electrophysiological correlates than just one early ERP component (Zion-Golumbic and Bentin, 2007). The vast majority of scientific studies nevertheless focus on the N170 and fewer studies report results on later components related to face perception along with early potentials. However, these studies indicate that profound processing of faces beyond basic sensory-perceptual properties is rather captured by late potentials (Ratcliff et al., 2009; Herbert et al., 2013; Bublitzky et al., 2014; Stolz et al., 2019). Consequently, the process of perceiving and interpreting a face under realistic conditions is not necessarily limited to the early components.

Due to the above-mentioned increase in spatial proximity, realism and thus self-relevance of persons when presented in VR, we expected more sophisticated and in-depth face processing under realistic conditions. This could already be evident in better discrimination between faces and controls under realistic conditions, as reflected by increased amplitude differences between stimulus types for the early components, particularly the N170. However, given the previously reported doubts about the face specificity of the N170, we rather

expected less discrimination between faces and controls under realistic conditions, reflected in absence of early amplitude differences between stimulus types in VR.

We instead hypothesized, based on the aforementioned results on later components in face processing, that a more realistic encounter with a person should lead to enhanced sensitivity processing of facial features, i.e., better discrimination between faces and controls, in VR, especially reflected by later components (Ratcliff et al., 2009; Herbert et al., 2013; Bublitzky et al., 2014; Schindler et al., 2017; Stolz et al., 2019).

Complementing the ERP analyses, we investigated the neural generators of realistic face perception and expected a larger network of neural structures involved in the underlying processes than under conventional conditions (e.g., Vuilleumier and Pourtois, 2007). We hypothesized that the fusiform gyrus, which is specialized in face perception and object recognition (McCarthy et al., 1997; Weiner and Zilles, 2016; Palejwala et al., 2021), and the inferior temporal gyrus, associated with higher visual processing and face individuation (Nestor et al., 2008; Pyles et al., 2013) would be part of the network. In addition, we expected to obtain sources related to higher cognitive functions indicating more profound face processing with regard to self-referential and emotional information under realistic conditions.

Since there are no results yet on a direct comparison of face perception under 2D and realistic conditions, our hypotheses concerning specific amplitude and topographical differences between the two modalities remain of overall exploratory nature.

2. Methods

2.1. Participants

Thirty participants were recruited from Osnabrück University. All participants were screened for psychological and neurological disorders and regular drug use. Only participants who met the inclusion criteria were eligible for the experiment. Additionally, previous experience with computer games and Virtual Reality and the recent usage of such media was documented. Participants had none to little experience with VR and did not wear a VR headset regularly within the last 4 weeks. If vision correction was necessary, only participants wearing contact lenses, not glasses, could participate. It was furthermore ensured that participants had not also been photographed for the stimulus creation (see 2.2) or knew any of the people whose pictures were presented to them (max. Recognition rate was below 7%, i.e., eight out of 120 faces). All participants gave informed written consent. Participants received either partial course credits or 15€ for their participation.

Four participants had to be excluded from participation and analyses due to unmet anamnesis criteria ($n=2$) or because they aborted the experiment ($n=2$). Ultimately, 26 data sets were selected for data analyses ($M_{\text{age}}=22.96$ years, $SD_{\text{age}}=3.1$ years, 20 female, 25 right-handed). The sample size is similar to other studies that conducted a VR experiment featuring face stimuli and also investigated the N170 as well as later components (Stolz et al., 2019).

2.2. Stimulus material

The stimulus material comprised 120 pictures of persons sitting on a stool in a plain living room as the background, and were rendered

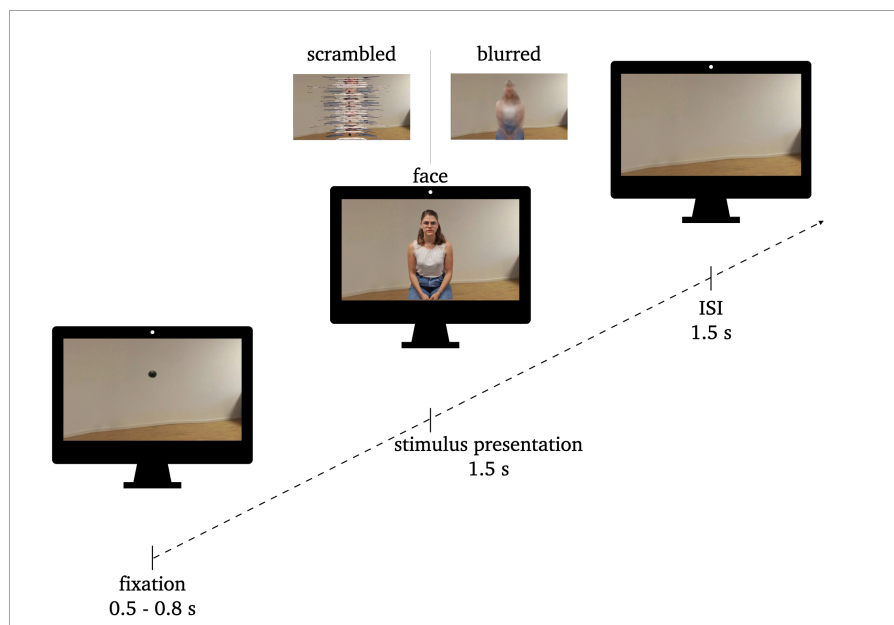


FIGURE 1

Procedure of stimulus presentation: 0.5–0.8s fixation, 1.5s stimulus presentation, 1.5s inter stimulus interval (ISI). Exemplary stimuli of face conditions and perceptual control conditions (scrambled, blurred) are illustrated.

as both, 2D and 3D-360° images. All of them showed neutral facial expressions. All images were recorded with the Insta360Pro VR-camera with 8k resolution at a distance of 62 cm to the person being photographed. The images were randomly assigned per participant to the two conditions (PC or VR), yielding in 60 individual images per condition, while ensuring that participants would not see the same person twice. In addition, two classical perceptual control pictures were presented, which are generally utilized in face processing research in order to control for the perceptual features of the stimuli (e.g., Herrmann et al., 2005a,b; Schwanger et al., 2009; Kuefner et al., 2010; Bombari et al., 2013; Rossion, 2014; Civile et al., 2018). The first control was a blurred version conserving color information and perceptual frame, i.e., stimulus size and shape, but without the same semantic relevance. The blurring was achieved by applying the glass-filter in Adobe Photoshop 2022 (Distortion filtering: Glass Filter, Distortion: 15, Smoothing: 1, Structure: Milk Glass, Scaling 200%). The second control was a scrambled version sustaining equivalent low-level perceptual visual properties (see Figure 1). For the scrambled pictures, the original pictures were cut in to stripes with a height of 10 pixels that were randomly rearranged in the vertical dimension.

The subtraction of the process of perceiving the blurred or scrambled image from the process of perceiving the normal person isolates the processes actually associated with person perception that are not influenced by the physical attributes of the stimuli like, e.g., size, color or spatial configuration. Controlling for these perceptual factors, i.e., having a perceptual baseline, thus allows for direct comparison of the processes across the different domains (PC and VR), so that the face specific processing reflected by ERP amplitude differences can be compared across modalities.

Taken together, each condition included 180 stimuli (60 normal, 60 blurred, 60 scrambled). The normal persons and their matching control images were always presented in the same modality, respectively.

2.3. Procedure

All participants completed both, the PC and the VR condition, while the order of both conditions was alternated between participants. Both conditions were conducted in the same soundproof and electrically shielded room suitable for Electroencephalographic (EEG) measurements. Thus, participants did not switch their location between conditions and were given a five-minute break to relax and get ready for the second condition. During this break, the EEG signal quality was checked. For both conditions, participants were instructed to passively watch the stimulus presentation and keep their movement to a minimum.

For the PC condition, participants were seated in front of a standard PC monitor (24", 1920 × 1200 resolution) with a constant distance of 115 cm to the screen, resulting in a horizontal viewing angle of 5° and a vertical viewing angle of 2.5°. The pictures were presented in 2D in the center of the screen with a size of 10 × 15 cm using Matlab for stimulus presentation.

For the VR condition, participants remained seated and were equipped with a VR headset (HTC Vive Pro 2, 2448 × 2448 pixel per eye, up to 120° field of vision, 120 Hz refresh rate). The pictures were presented in 3D-360° in real-life size at a distance of 62 cm (horizontal viewing angle: 98°; vertical viewing angle: 42°) via the video-game engine Unity 5 (Version 2020). Triggers were sent by Unity and synchronized using Lab Streaming Layer for Unity (LSL by SCCN).¹

Each of the 180 trials of each condition followed the same sequencing (see Figure 1). The pictures were presented for 1.5 s. They were preceded by a fixation dot (0.5–0.8 s) and followed by an interstimulus interval (ISI; background image without person; 1.5 s).

¹ <https://github.com/sccn/labstreaminglayer>

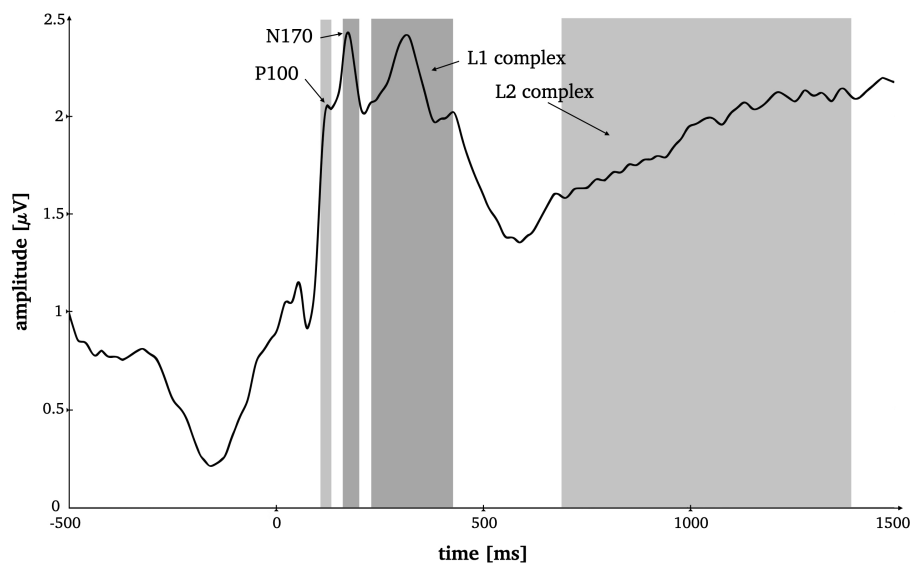


FIGURE 2

Time-by-amplitude plot of the root mean squared ERP averaged over all electrodes for the selection of appropriate time windows for all ERP components. Grey highlighted sections mark the time windows for P1 (95–125ms), N170 (165–195ms), L1 (230–420ms) and L2 (685–1,385ms).

The participants were instructed to blink or move only during the ISI, i.e., while the room they saw was empty. Each trial lasted between 3.5 to 3.8 s resulting in a total run time of approximately eleven minutes per condition. The sequence of the stimulus presentation was the same for both conditions, the only difference was the modality (PC vs. VR). In VR, the participants remained in the same environment (i.e., the living room) for the entire time. The fixation dot and the stimuli appeared in front of them in the respective sequence without any changes to their surroundings. For the ISI the room stayed empty.

Due to the sensitivity of EEG data to motion-induced artifacts, participants were asked to keep motion to a minimum and refrain from looking around in the VR environment. The feasibility of EEG measurements while additionally wearing a head-mounted display (HDM) has been investigated before with special concern for EEG signal quality. However, EEG signal quality was not influenced negatively and the combination of VR using HDM and EEG was rated feasible (Hertweck et al., 2019; Tauscher et al., 2019).

2.4. Electrophysiological recordings and preprocessing

An electroencephalogram (EEG) with 128 electrodes, attached in accordance with the international 10-20-system was recorded for the duration of the whole experimental procedure (PC and VR condition). The Active-Two amplifier system from BioSemi (Amsterdam, Netherlands) was used. The sampling rate was 512 Hz, the bandwidth (3 dB) 104 Hz. Additionally, horizontal electrooculogram (hEOG) and vertical electrooculogram (vEOG) were recorded and a common mode sense (CMS) and a driven right leg (DRL) electrode were applied. During the PC condition, the EEG was recorded on the investigators' computer using ActiView702 Lores. For the VR

condition, the trigger stream from Unity was transmitted to Lab Streaming Layer to synchronize the EEG data stream and Unity triggers.

The first preprocessing step necessary only for the VR-condition comprised the merging of the EEG data stream and trigger stream via the EEGLAB add-on MoBi-Lab (Ojeda et al., 2014). All further preprocessing steps were applied to the recordings of both modalities using EEGLAB (Delorme and Makeig, 2004).

The data were re-referenced to average reference, high-pass filtered at 0.25 Hz and low-pass filtered at 30 Hz. Bad channels were identified using the automatic channel removal add-on (ASR; Mullen et al., 2015) and interpolated. All channels were linearly detrended for elimination of extended potential drifts. Artifact rejection was applied using independent component analysis (ICA; Delorme et al., 2007). Specifically, an automatic automated ICA component labeling was performed (ICLabel v1.4; Pion-Tonachini et al., 2019; artifact selection confidence of 90% for Muscle, Heart, Line Noise und Channel Noise and 80% for Eye Artifacts). The results of the ICA were visually verified. For epoching, the time window around the trigger onset was set from −500 to 1,500 ms and the baseline correction was set from −300 to 0 ms before trigger onset. Per modality and within modality per stimulus category (face, blurred, scrambled), grand means were computed resulting in six individual ERPs (i.e., VR-face, VR-blurred, VR-scrambled, PC-face, PC-blurred, PC-scrambled).

2.5. ERP components in electrode space

The time windows and electrode sites for the classical ERP components were selected based on prior literature (Latinus and Taylor, 2006; Rossion and Jacques, 2008; Boehm et al., 2011; Dering et al., 2011), as well as visual inspection of the root-mean-squared ERP (see Figure 2) and the mean topographies across modality and stimulus type (see Figure 3) of all conditions. The

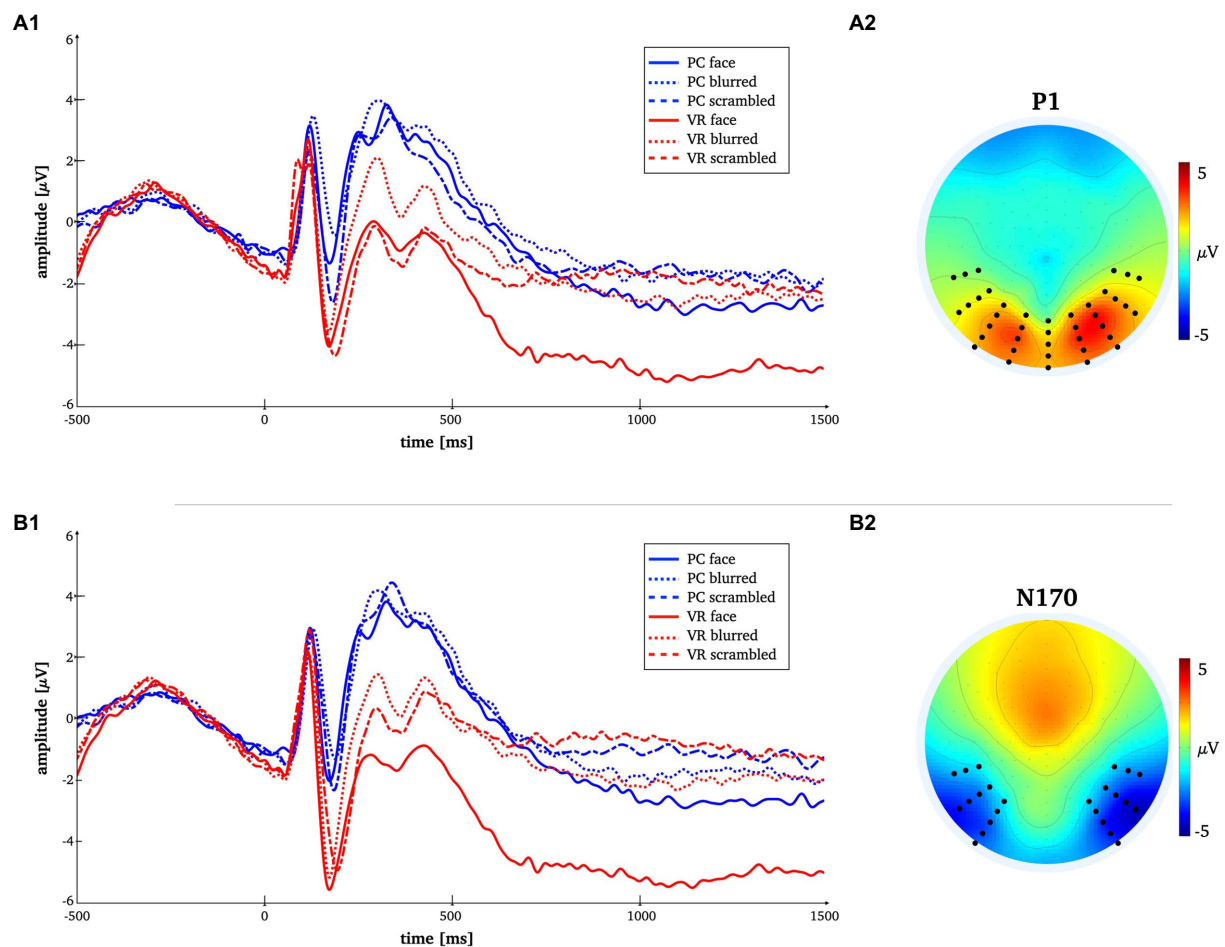


FIGURE 3

Time-by-amplitude plot of the mean P1 and N170 amplitudes for all conditions [panels **A1,B1**]. Mean topographies across conditions used for ERP averaging [panel **A2,B2**]. The electrodes selected for analyses are indicated. For the P1 electrodes Oz, O1, O2, P7, PO7, P8, PO8, TP7, TP8 and those in close vicinity were used. For the N170 electrodes P7, P8, PO7, PO8, P10, P9, PO10, PO9, TP7, TP8 and those in close vicinity were used.

P1 was analyzed at posterior midline and lateral (i.e., Oz, O1, O2) as well as occipito-parietal (i.e., PO7, PO8, P7, P8) electrodes in a 95–125 ms time window. The N170 was analyzed at parietooccipital electrodes (i.e., P7, P8, P9, P10, PO7, PO8, PO9, PO10). Here, the time window was set from 165 to 195 ms. The P1 and the N170 amplitudes were computed by calculating the mean voltage across the selected electrodes and time windows, respectively.

Two more consecutive potential complexes were selected for analysis, both analyzed at all 128 electrode positions. Prior literature on late potentials in face and object perception served as basis for our hypotheses on time window selection (Bublitzky et al., 2014; Stolz et al., 2019; Johnsdorf et al., 2023), which was ultimately decided upon based visual inspection of the root mean squared ERP (Figure 2). The L1 component was set from 230 to 420 ms around the local maximum approximately corresponding to the known early posterior negativity potential (EPN; Bublitzky et al., 2014). The L2 was set from 685 to 1,385 ms which comprises a late local minimum and the consecutive amplitude increase (see Figure 2) approximately corresponding to the known late positive potential (LPP; Bublitzky et al., 2014).

2.6. Statistical analysis

All data preparation for statistical testing, i.e., generating respective means for selected time windows and electrodes, as well as the correlation analyses was implemented in Matlab (Version R2021b). The rmANOVA was done via IBM SPSS Statistics (Version 27), while *post hoc t*-testing was additionally double-checked in Matlab. For additional robust statistical testing the statistic software R (Version 4.2.1) was used.

2.6.1. P1 and N170

The EEG data for P1 and N170 were analyzed using a 2×3 repeated-measurements ANOVA (rmANOVA) with the within-subject factors “modality” (VR vs. PC) and “stimulus type” (face vs. blurred vs. scrambled). Whenever necessary, Greenhouse–Geisser-corrected *p*-values are reported. Significant effects of rmANOVA were complemented by *post hoc t*-tests within each modality, as well as for between-modality interactions. For the N170, we complement our analysis with a robust rmANOVA employing the same factors as well as robust *post hoc t*-tests (both without bootstrapping). For the robust rmANOVA, we used the *wwtrim* function from the WRS package

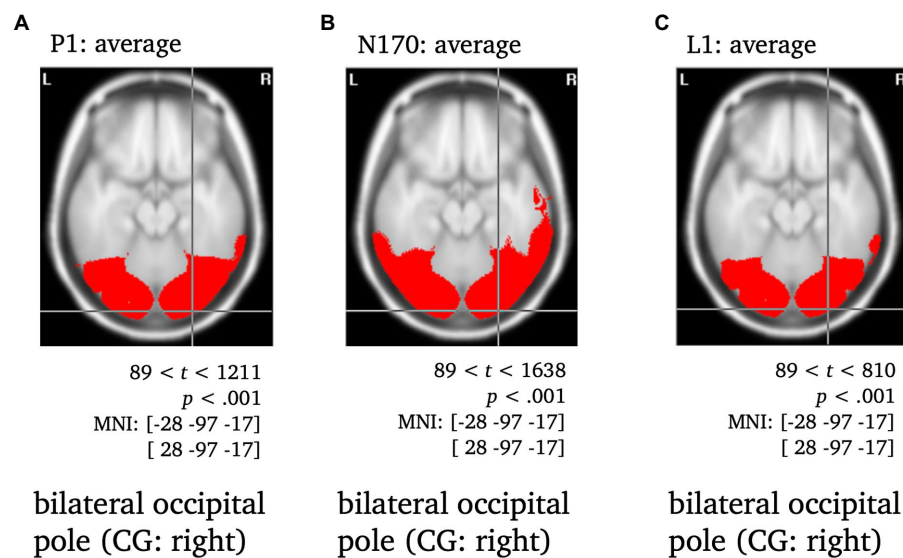


FIGURE 4

Statistically significant sources for the time span of the P1, N170 and L1 averaged across conditions. Statistically significant differences in activity are marked red, with $p < 0.001$, $t_{crit} = 89$. Per panel, the center of gravity (CG) is labeled and the respective MNI coordinates are given.

with default parameters, i.e., trim = 0.2, and the yuend function from the WRS2 package for the t -tests (Mair and Wilcox, 2020). Robust statistics can be used to ensure replicability with small sample sizes.

2.6.2. L1 and L2

For L1 and L2, the EEG data were analyzed using 2D-correlation (see formula below) as a first step to determine whether similarities between the topographies of the stimulus types were comparable in both conditions:

$$r_{2D} = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \bar{A})^2)(\sum_m \sum_n (B_{mn} - \bar{B})^2)}}$$

Additionally, Pearson correlation coefficients across all 128 electrodes were calculated per participant for all within-modality comparisons and also for relevant between-modality comparisons. The mean correlation coefficient and the number of significant correlations was determined for all comparisons. To test statistically whether similarities between stimulus types varied between modalities, the within-modality comparisons were then t -tested against respective pairings of the other modality (i.e., PC face ~ PC blurred vs. VR face ~ VR blurred).

2.7. ERP components in source space

To determine the differences in activation of the cortical generators involved in face perception under conventional laboratory conditions compared to realistic conditions, variable resolution electromagnetic tomography (VARETA; Bosch-Bayard et al., 2001) was applied. VARETA provides an intracranial distribution of current densities in source space that is spatially smoothest and highly

compatible with the amplitude distribution in electrode space (Gruber et al., 2006; Martens et al., 2011). We applied an inverse solution that comprised 3,244 grid points arranged in a 3D grid. The grid was defined by a Leadfield matrix and corresponded to the placement of the 128-channel EEG-system (10-20-system).

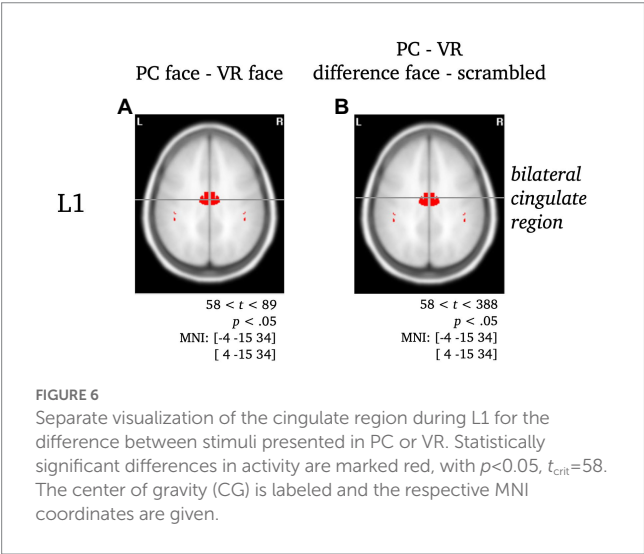
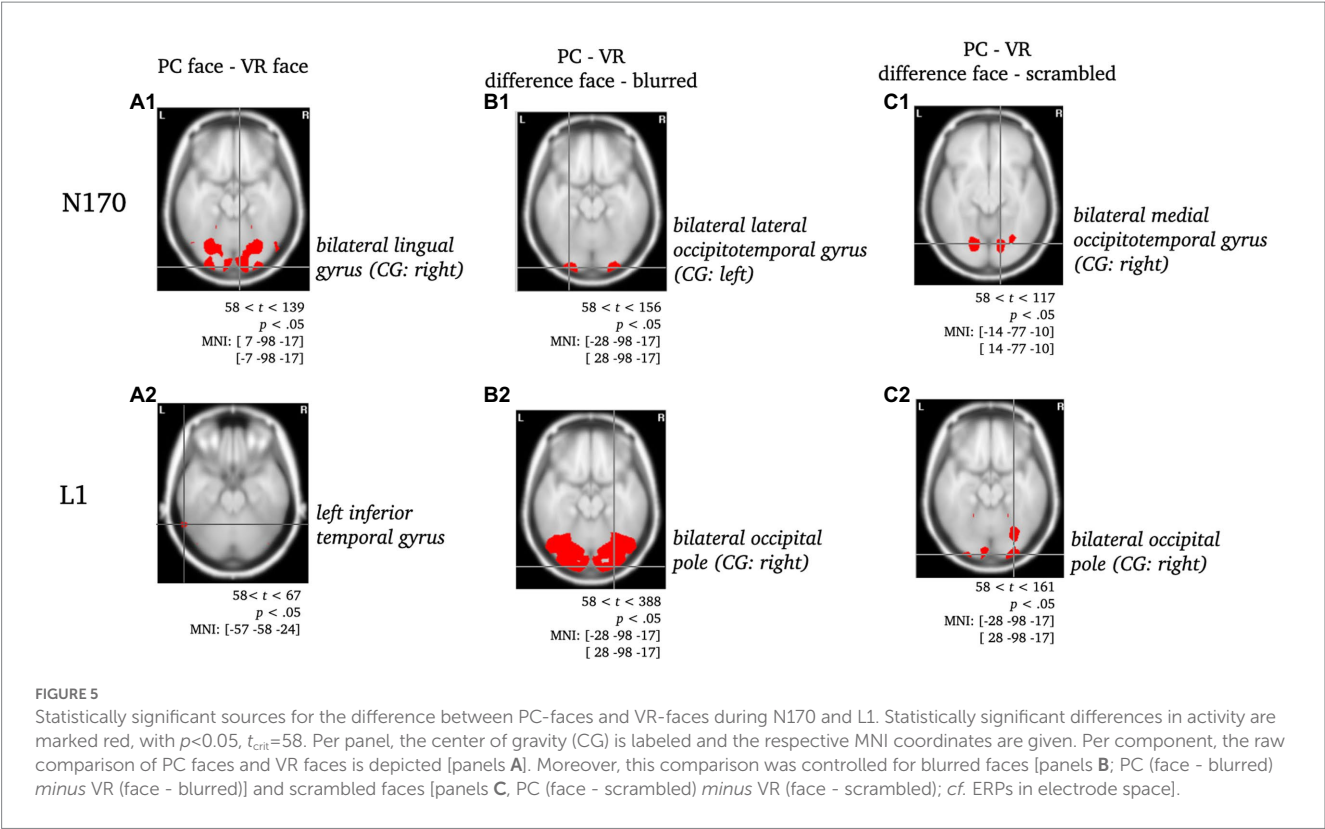
To localize differences in activation patterns, Hotellings T^2 -test was performed per effect of interest (see ERP components in electrode space). As a first validation step, the sources of the P1, N170 and L1 components of the ERP were localized to validate the use of VARETA with the current data set (see Figure 4). For all t -tests regarding the average across all conditions, the critical t -value was $t_{crit} = 89$ with a significance level of $p < 0.001$. After a consistency check against previous publications (e.g., Di Russo et al., 2002; Gruber et al., 2006), the sources of further effects of interest were examined. For these comparisons between conditions (see N170, L1; Figures 5, 6) the significance level was set to $p < 0.05$ and the critical t -value was $t_{crit} = 58$. Significant voxels were projected onto the cortical surface which was constructed on the basis of the average probabilistic MRI brain atlas by the Montreal Neurological Institute (MNI; Evans et al., 1993). The brain region's names for significant voxels were identified by the brain electrical tomography (BET) Neuronic Tomographic viewer.

3. Results

3.1. ERP components in electrode space

3.1.1. P1

The rmANOVA for the P1 component revealed no significant main effects for the factors “modality” and “stimulus type” or the interaction of both ($F_{modality}(1, 25) = 0.84$, $p = 0.37$; $F_{stimulus}(2, 50) = 2.7$, $p = 0.09$, $\epsilon = 0.78$; $F_{interaction}(2, 50) = 0.29$, $p = 0.75$). The respective descriptive statistics are given in Table 1 and Figure 7.



3.1.2. N170

The rmANOVA for the N170 component revealed significant main effects for the factor “modality” ($F_{modality}(1, 25) = 49.27$, $p < 0.001$, $\eta^2 = 0.66$), but not for the factor “stimulus type” ($F_{stimulus}(2, 50) = 1.85$, $p = 0.178$, $\epsilon = 0.75$), while the interaction of “modality” and “stimulus type” was significant ($F_{interaction}(2, 50) = 3.87$, $p = 0.029$, $\eta^2 = 0.25$). The respective descriptive statistics are given in Table 1 and the results of the robust ANOVA are given in Table 2.

A significantly more negative N170 amplitude was found for normal compared to blurred persons [$t(25) = -2.08$, $p = 0.048$,

TABLE 1 P1 and N170 mean amplitudes, standard deviations and confidence intervals for both modalities and all stimulus types.

	<i>M</i>	<i>SD</i>	Confidence interval	
			Lower limit	Upper limit
P1 – PC				
Face	2.73	2.0	1.93	3.53
Blurred	2.63	1.8	1.92	3.35
Scrambled	1.95	2.6	0.9	3.0
P1 – VR				
Face	2.19	3.5	0.77	3.61
Blurred	1.97	2.8	0.85	3.1
Scrambled	1.59	4.5	−0.22	3.41
N170 – PC				
Face	−1.48	2.9	−2.63	−0.32
Blurred	−0.87	3.0	−2.1	0.4
Scrambled	−0.183	3.4	−3.2	−0.5
N170 – VR				
Face	−5.11	4.0	−6.72	−3.5
Blurred	−4.39	3.8	−5.94	−2.85
Scrambled	−4.28	4.4	−6.05	−2.52

$d = -0.41$] and for scrambled compared to blurred persons [$t(25) = 2.94$, $p = 0.007$, $d = 0.58$] in the PC modality, but no significant difference in amplitude for normal compared to scrambled persons

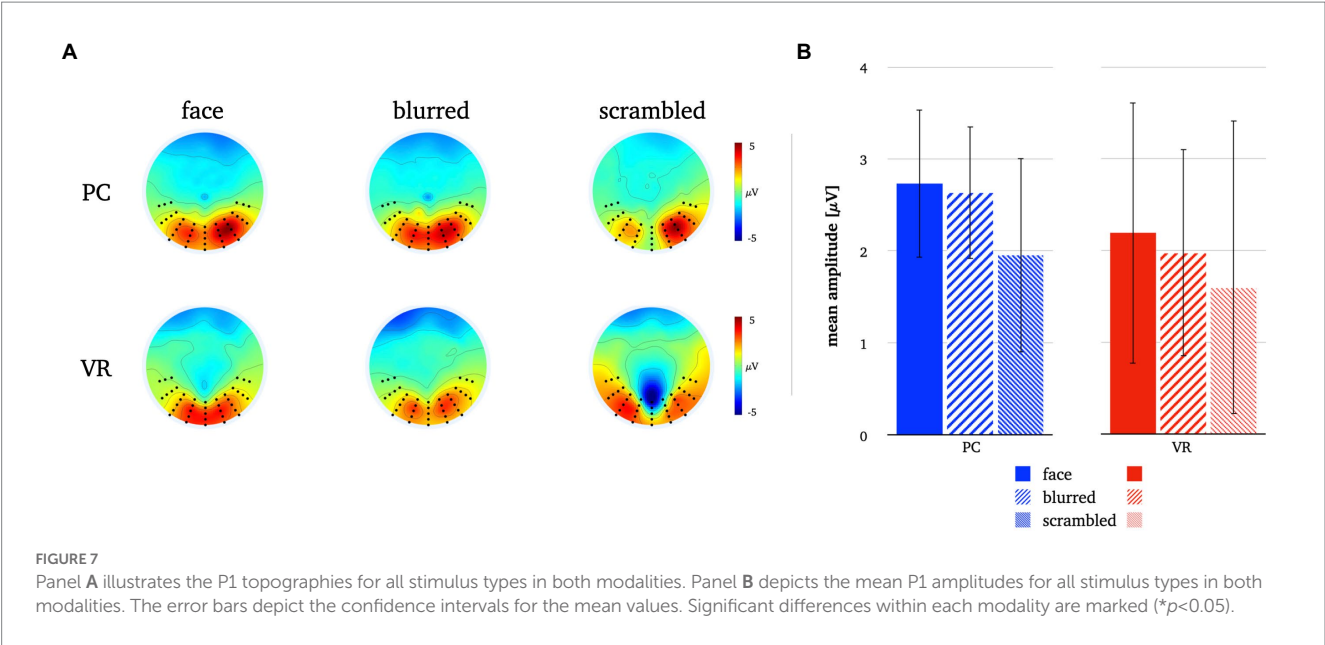


TABLE 2 Results of 2×3 repeated-measurements ANOVA (rmANOVA) with the within-subject factors “modality” (VR vs. PC) and “stimulus type” (face vs. blurred vs. scrambled).

		df	F	p	partial η^2
Modality	Standard	1	49.27	<0.001***	0.66
	Robust	2	57.79	<0.001***	-
Stimulus type	Standard	2	1.85	0.167	0.07
	Robust	1	3.03	0.05*	-
Interaction	Standard	2	3.87	0.027*	0.13
	Robust	2	0.29	0.075	-

For every comparison, robust *F*-statistics are given in the second line. Significant effects are marked (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$).

[$t(25) = -0.08$, $p = 0.434$]. In VR, there were no significant N170 amplitude differences between stimulus types (all $ps > 0.05$, see also Figure 8; Table 3).

Comparing stimulus types across modalities revealed significantly more negative N170 amplitudes for all stimulus types in VR when compared to PC (see Table 3). Lastly, the analyses of interaction effects showed significantly more negative N170 amplitude for normal faces and blurred faces in VR when priorly subtracted by the scrambled type (see Table 3; Figure 9). For the results of the robust *t*-tests for all comparisons within and between modalities see Table 3.

3.1.3. L1

A two-dimensional correlation as well as mean correlation coefficients for comparison between topographies within each modality revealed higher similarity between stimulus types within the PC condition. The results for the direct comparison of stimulus types across modalities also indicate low similarity. For detailed statistics please refer to Figure 10.

Pairwise *t*-tests of the correlation coefficients for stimulus comparisons between modalities showed strong differences between PC and VR. The similarity between the stimulus types is significantly

lower within VR compared to PC, which is furthermore represented by a smaller number of significant correlations.

3.1.4. L2

The two-dimensional correlation as well as mean correlation coefficients for comparison between topographies within each modality disclosed bidirectional differences between stimulus types for PC and VR. The results for the direct comparison of stimulus types across modality indicate moderate similarity. For detailed statistics please refer to Figure 11.

Pairwise *t*-tests revealed moderate differences for normal compared to scrambled persons between VR and PC, but no significant differences for normal compared to blurred persons or blurred compared to scrambled persons were found.

3.2. ERP components in source space

3.2.1. P1, N170 and L1 across modality and stimulus type

All components were localized to the bilateral occipital pole, with the center of gravity in the right hemisphere when averaged across conditions and stimulus types and were thus consistent with previous literature (e.g., Di Russo et al., 2002; Gruber et al., 2006; Figure 4).

N170: Regarding the N170 component, the difference between PC faces and VR faces was localized to the bilateral lingual gyrus with the center of gravity in the right hemisphere (see Figure 5A1). When controlling for blurred faces, the difference was localized to the bilateral lateral occipitotemporal gyrus (CG: left; see Figure 5B1), whereas it localized to the bilateral medial occipitotemporal gyrus (CG: left; see Figure 5C1) when controlling for scrambled faces.

L1: The difference between PC faces and VR faces yielded significantly different activity in the left inferior temporal gyrus regarding the L1 component (see Figure 5A2). When controlling for

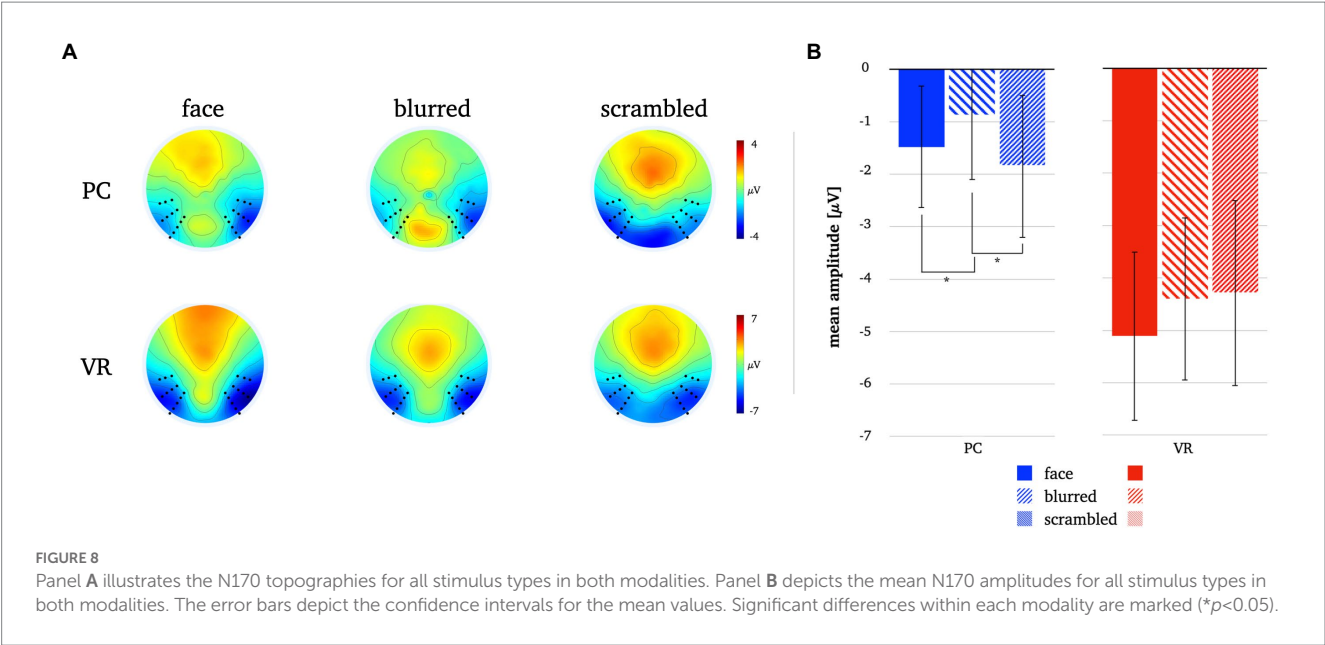


TABLE 3 Pairwise comparisons of N170 amplitudes within and between modalities.

		<i>df</i>	<i>t</i>	<i>p</i>	Cohen's <i>d</i>
PC					
Face – Blurred	Standard	25	−2.08	0.048*	−0.41
	Robust	15	−2.64	0.019*	0.18
Face – Scrambled	Standard	25	0.08	0.434	0.16
	Robust	15	−0.53	0.605	0.06
Blurred – Scrambled	Standard	25	2.94	0.007**	0.58
	Robust	15	1.31	0.21	0.11
VR					
Face – Blurred	Standard	25	−1.83	0.08	−0.36
	Robust	15	−1.611	0.128	0.15
Face – Scrambled	Standard	25	−1.48	0.151	−0.29
	Robust	15	−1.01	0.326	0.12
Blurred – Scrambled	Standard	25	−0.24	0.813	−0.05
	Robust	15	0.15	0.885	0.01
PC versus VR					
Face	Standard	25	6.17	<0.001***	1.21
	Robust	15	5.22	<0.001***	0.62
Blurred	Standard	25	4.35	<0.001***	1.72
	Robust	15	8.89	<0.001***	0.61
Scrambled	Standard	25	3.65	<0.001***	0.83
	Robust	15	5.25	<0.001***	0.5
Face – Blurred	Standard	25	0.26	0.797	0.05
	Robust	15	−0.32	0.752	0.05
Face – Scrambled	Standard	25	2.37	0.026*	0.47
	Robust	15	1.45	0.167	0.25
Blurred – Scrambled	Standard	25	2.05	0.035*	0.44
	Robust	15	2.03	0.061	0.36

For every comparison, robust *t*-statistics are given in the second line. Significant differences are marked (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$).

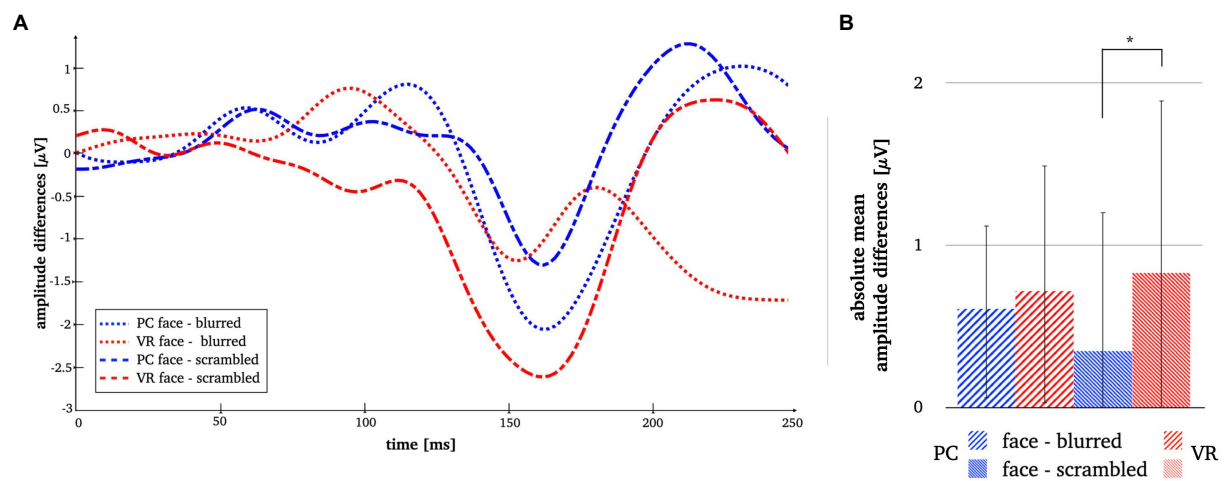


FIGURE 9

Panel A illustrates the N170 amplitude differences after subtraction of the perceptual controls in both modalities. Panel B depicts the absolute mean face amplitudes after subtraction of the perceptual controls in both modalities. The error bars depict the confidence intervals for the mean values. Significant differences are marked (* $p < 0.05$).

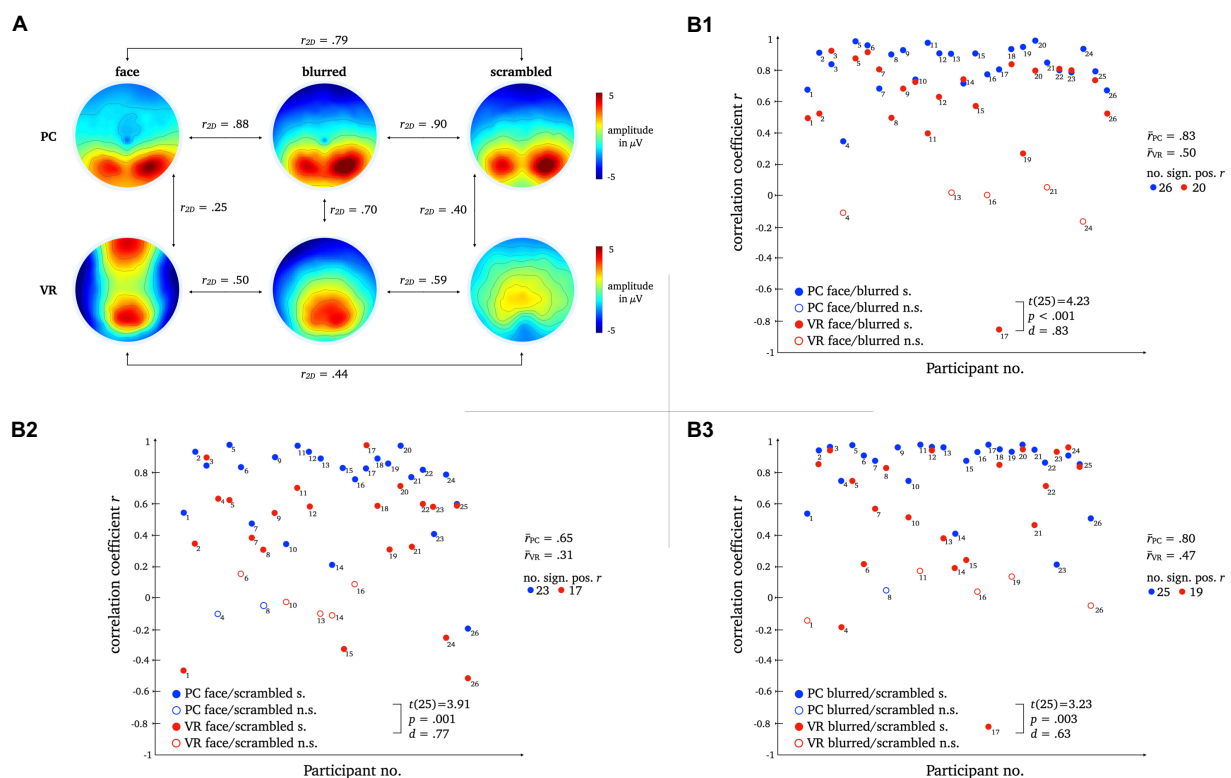


FIGURE 10

L1 topographies with 2D-correlation coefficient r for within and between modality comparisons [panel A]. The scatter plots illustrate individual correlation coefficients for stimulus type comparisons between modalities: Face vs. Blurred [panel B1], Face vs. Scrambled [panel B2] and Blurred vs. Scrambled [panel B3]. T -Test statistics, mean correlation coefficients and number of significant correlation coefficients are given.

either blurred or scrambled faces, the differences in activity localized to the bilateral occipital pole (see Figures 5B2,C2). Interestingly, the difference between PC faces and VR faces was accompanied by

significant differences in the activity of the cingulate region for the raw difference as well as when controlling for scrambled faces, but not when controlling for blurred faces (see Figure 6).

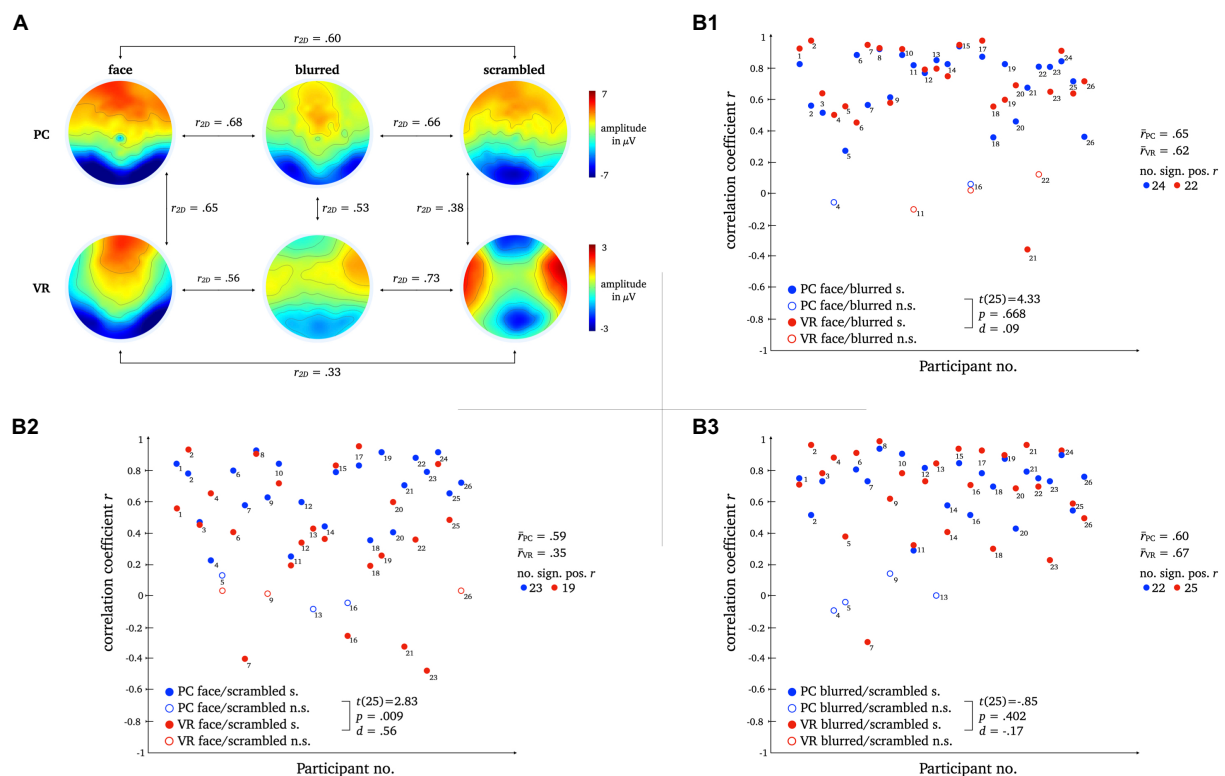


FIGURE 11

L2 topographies with 2D-correlation coefficient r for within and between modality comparisons [panel A]. The scatter plots illustrate individual correlation coefficients for stimulus type comparisons between modalities: Face vs. Blurred [panel B1], Face vs. Scrambled [panel B2] and Blurred vs. Scrambled [panel B3]. T -Test statistics, mean correlation coefficients and number of significant correlation coefficients are given.

4. Discussion

The aim of this study was to investigate the neuronal mechanisms of realistic human face processing by translating the conventional laboratory setup into a realistic setting using VR. To take a first step towards bridging the gap between classical laboratory designs and reality, we followed the overall rationality of laboratory conventions, both in terms of experimental setup as well as the analytical methods, and thereby maintained comparability of the electrophysiological results.

To this end, randomized picture sequences of persons and two perceptual controls were presented to the participants in a blocked within-design (see Methods) under realistic conditions *via* a head-mounted display (VR modality) and under conventional 2D conditions *via* a PC monitor (PC modality). In our VR condition, participants were confronted with three-dimensional, real-life-sized persons sitting directly in front of them, undercutting social distance and invading the personal space. Due to the spatial proximity of the VR stimuli, a feeling of social involvement is created thus requiring processing of complex contextual and self-relevant information. To control for low-level perceptual differences (e.g., size and shape) between the two presentation modalities, affecting cognitive processing, the across modality comparisons were carried out after subtracting the amplitudes of the control images (i.e., blurred and scrambled pictures). This allows for comparing the N170 amplitudes between PC and realistic conditions beyond basal perceptual processes

of the stimuli's low-level visual features (e.g., size, shape, color). The control images thus served as a perceptual baseline. Albeit reporting results from robust statistical methods, in line with existing laboratory research, we base our interpretations on the conventional t -test.

By investigating classical ERP components commonly related to face processing, i.e., N170 (e.g., Itier and Taylor, 2004; Blau et al., 2007; Rossion and Jacques, 2011) and P1 (Itier and Taylor, 2002; Herrmann et al., 2005a; Thierry et al., 2007; Dering et al., 2009; Kuefner et al., 2010), and later components (L1, L2) as well as their cortical generators, we compared face processing mechanisms between conventional laboratory and realistic conditions by means of VR. Most importantly, our results under laboratory conditions generally replicate previous studies showing no indication of a face sensitivity of the P1 (e.g., Jemel et al., 2003; Ganis et al., 2012), a face specificity of the N170 when presented on a PC monitor (e.g., Eimer, 2000; Jemel et al., 2003; Rousselet et al., 2008; Rossion, 2014; Civile et al., 2018) and the relevance of later components in perceptual processing (e.g., Herbert et al., 2013; Bublatzky et al., 2014; Schindler et al., 2017). Although the ERPs obtained under laboratory conditions exhibited the same topographies at early stages of face processing, at later stages they differed markedly from their VR counterparts. Especially the L1 exhibited superior discrimination between faces and controls only in VR. These results indicate a more fine-tuned processing of faces under realistic conditions in VR, casting doubt on the general meaningfulness of the N170 as a singular marker for real-life face processing.

4.1. ERP components in electrode space

4.1.1. P1

Investigations concerning a potential face sensitivity of the P1 component yielded no significant main effects, i.e., no discrimination between stimulus types in either modality.

The P1 seems to be essentially sensitive to certain stimulus properties but not specifically for faces, when presented on a PC monitor or a VR headset. In line with previous study results, the P1's stimulus sensitivity presents a rather inconclusive picture. Whether the P1 is fundamentally insensitive for faces (see, e.g., [Ganis et al., 2012](#)), sensitive for object category (see, e.g., [Thierry et al., 2007](#); [Dering et al., 2009](#)) or even task-sensitive (see, e.g., [Dering et al., 2009](#)) is unclear.

Under realistic conditions, the P1 did not discriminate between any stimulus types and thus showed no face sensitivity at all. These results extend previous research and suggest that the P1 is not a suitable neural correlate for the perception and processing of faces under realistic conditions.

4.1.2. N170

Most importantly, confirming previous results on the N170 component face effect, we replicated other laboratory studies finding stronger amplitude deflections for 2D faces compared to non-face perceptual controls (e.g., [Eimer, 2000](#); [Jemel et al., 2003](#); [Rousselet et al., 2008](#); [Rossion, 2014](#)). Furthermore, we extended results from other experimental setups using 3D-presentations that obtained the N170 using VR ([Stolz et al., 2019](#)) by comparing it directly to conventional laboratory conditions.

The comparison of the N170 amplitudes within each modality revealed no significant differences between faces and controls within the VR modality. Even though the *p*-values of the amplitude differences under realistic conditions reached trend level, no evidence for the discrimination of stimulus types by the N170 could be found. Thus, the N170 loses some of its discriminatory power when obtained under more realistic conditions with VR.

When contrasting the N170 amplitudes of the three stimulus types directly across modalities, VR leads to more negative amplitudes than PC. Considering the comparably larger stimulus size, the area occupied on the retina and the retinotopic organization of the visual cortices, respectively, these amplitude differences are most likely result for the stimuli's physical perceptual features ([Busch et al., 2004](#); [De Cesare and Codispoti, 2006](#); [Josef Golubic et al., 2011](#); [Pfabigan et al., 2015](#)). Previous research has shown that the perception of the human body independent from the face, also elicits a typical N170 response, in some cases with a delayed latency, i.e., a N190 response ([Stekelenburg and De Gelder, 2004](#); [Thierry et al., 2006](#)). In both conditions of this study, we deliberately used stimuli that comprised the face as well as upper body of persons as one would encounter it in real life as well. The amplitudes are therefore influenced by the perception of the upper body as well, but this affects both conditions which is why it is unlikely to explain the amplitude differences between the modalities.

The interaction effects across modality, i.e., subtracting the amplitudes of the control pictures before comparing the N170 amplitude, revealed stronger N170 deflection under realistic conditions for one of the perceptual baselines. Controlling for perceptual frame and color information by subtracting the amplitude of the blurred

image, led to comparable N170 amplitudes in both modalities. The subtraction of the amplitude of the scrambled image, i.e., low-level perceptual visual features, resulted in a stronger N170 deflection for VR faces. However, the amplitude difference face *minus* scrambled was not significantly different from zero, providing further evidence against the specificity for faces under realistic conditions. Hence, the modality effect actually results from smaller variances within the 2D modality, showing that the N170 face specificity can be replicated in the conventional laboratory but not under realistic conditions.

Taken together, differences between the N170 obtained under realistic conditions as opposed to conventional laboratory conditions are apparent but relatively small, which is reflected by equally small effect sizes (see Results). Our results confirm the long hold notion that the N170 specifically reflects cognitive processes related to face perception under conventional 2D PC conditions, while our study does not provide any evidence that under realistic conditions the N170 likewise indicates face perception as it does not differentiate between different types of stimuli. Thus, our data shows that the N170 specificity seems to be a domain or modality related effect ([Schöne, 2022](#)). The comparison of perceptual-baseline-corrected amplitudes furthermore showed that the amplitude variations between stimulus types are rooted in the PC effects. It has yet to be determined what might be the crucial factor diminishing the meaningfulness of N170 under realistic conditions. So far, one can only speculate as this study only took a first step in that direction. However, as described in the introduction, considering the much more immersive character of a visual scene in VR in which a presented person is embedded and the spatial proximity with which it can be done, it seems likely that the encounter with a human face under such conditions – as it would occur in real life – requires complex cognitive processing that differs from what we find when people watch faces on computer screens.

Previous studies have shown that the face specificity of the N170 cannot be replicated under all conditions, even in the conventional laboratory. For example, the N170 amplitude has been shown to also be sensitive to factors such as inter-stimulus perceptual variance (ISPV; [Thierry et al., 2007](#); [Dering et al., 2009](#)). In these studies, ISPV significantly modulates the N170 amplitude while the object category (i.e., face or car) does not. Moreover, cars produced stronger N170 deflections than faces ([Dering et al., 2009](#)), raising doubt regarding the N170 face effect. Results on the inversion effect on face perception obtained in a discrimination task using faces and cars show that the sensitivity of the N170 could also be explained by topographical differences and stimulus specific neural generators ([Boehm et al., 2011](#)). Most importantly, the N170 does not reflect behavioral improvements in social functions ([Key and Corbett, 2020](#)), which further questions whether the N170 reliably indexes face perception to be generally applicable even outside the conventional laboratory. Reflected in the results of this study, the N170's face specificity is only obtainable for faces that were presented on a PC monitor, but not consistent when transferred to realistic conditions.

4.1.3. L1 and L2

In contrast to the results of the N170 component, the investigation of the late components revealed the opposite picture. The correlation analyses yielded higher similarity between topographies of the three

stimulus types within PC as opposed to VR. The PC topographies are very similar suggesting that the same object perception mechanism is used for faces and controls, i.e., very different stimulus types. In contrast, the VR topographies differ considerably, implying distinct neural mechanisms for perception of faces, silhouettes and objects. Moreover, the topographies for each stimulus type differ between PC and VR, further supporting the stimulus specificity of the neural mechanisms applied under realistic conditions. The L1 component clearly differentiates more effectively between stimulus types under realistic conditions.

The results for L2 tie in well with the results for L1. The PC topographies are very similar as well, while the VR topographies are much more distinct. Across modalities, the topographies for faces and blurred controls look moderately similar, but for scrambled controls still markedly different. Again, the increased differentiation between face and object perception due to greater topographic differences under realistic conditions suggests face-specific neural mechanisms operating when encountering a realistic face that are not required for the monitor. Hence, as initially considered, examining face perception as an isolated process using typical face stimuli that are outside an egocentric reference frame and devoid of social context, initiates domain specific neural mechanisms that do not possess the same functional properties as those required for real-life face processing.

Extending previous laboratory studies, later components reflect said mechanisms of realistic face processing. In contrast to earlier components, later potentials are linearly related to stimulus realism (Schindler et al., 2017), modulated by socially relevant emotional expressions and affective contexts (Bublitzky et al., 2014; Stolz et al., 2019) and especially sensitive for self-related emotions (Herbert et al., 2013). Processing of actually self-relevant emotional and contextual information, such as, e.g., threat towards oneself, seems to not be captured by the N170 component. Thus, consistent with laboratory results, late components discriminate faces and controls under realistic conditions, as they exhibiting much more discriminatory potential than the N170.

4.2. ERP components in source space

4.2.1. N170

The source analysis of the N170 resulted in modality differences in the lateral and medial occipitotemporal gyrus revealing differing activation under VR and PC conditions. The medial occipitotemporal gyrus, which comprises the lingual gyrus, the parahippocampal gyrus, and the lateral occipitotemporal gyrus, which is also known as the fusiform gyrus, are functionally connected and involved in higher-order visual processing (McCarthy et al., 1997; Weiner and Zilles, 2016; Palejwala et al., 2021; Williams, 2021). Especially the fusiform gyrus is specialized in face perception and object recognition (McCarthy et al., 1997; Weiner and Zilles, 2016). The lingual gyrus, linking fusiform and parahippocampal gyrus, is related to processing of complex visual stimuli and their basic characteristics, such as emotional facial expressions, and moreover provides access to visual memory storage (Kozlovskiy et al., 2014; Palejwala et al., 2021). The parahippocampal gyrus is associated with a neural network processing contextual associations (Aminoff et al., 2013) and related to assessment of spatial configurations of objects while not

determining object identity (Bohbot et al., 2015). With regard to the results at hand, the obtained neural generators of the N170 suggest a first basic processing of faces on a primarily sensory-perceptual level that allows recognition of the stimulus being a face under both, laboratory and realistic conditions. However, taking into account the ERP results, a N170 face specificity is evident only under laboratory conditions, whereas under realistic conditions, only a low-level sensitivity is observed.

4.2.2. L1

For the source analysis of the L1 component, modality differences emerged in the left inferior temporal gyrus and cingulate region. The inferior temporal gyrus constitutes a higher level of visual processing, merging various higher cognitive functions such as visual perception, emotion regulation and memory (Miyashita, 1993) and is moreover related to person-specific semantic knowledge (Giovanello et al., 2003) as well as face individuation (Nestor et al., 2008; Pyles et al., 2013). As a core midline structure, the cingulate region is part of a network responsible for self-referential information processing (Northoff et al., 2006) and due to its functional connection to the hippocampus and the amygdala, it forms an important connection hub playing a role in long-term memory processing of emotional relevance of stimuli (Bubb et al., 2017; Rolls, 2019), such as familiar faces (Pierce et al., 2004). In contrast to the N170, the neural generators of the L1 reveal involvement of complex face-specific cognitive functions, such as memory and emotion regulation, suggesting more in-depth processing of the presented faces, i.e., consulting self-referential information and recognition of familiar faces. These assumptions are in line with aforementioned ERP findings in face perception, showing late components to be modulated by contextual, emotional and self-related information.

Taken together, contrasting neural sources under conventional as opposed to realistic conditions revealed an overall picture of face perception that would be expected when investigating realistic, self-relevant face processing. Initially, still within the time course of the N170, a face is detected as such, including recognition of its spatial configuration and contextual associations which is further supported by access to visual memory storage. However, processing of emotional relevance and retrieval of self-relevant face-specific information transpires on a broader time scale, reflected by the L1 component. Here, a face is recognized as an individual object with an identity and actual relevance within the observer's self-reference frame. Thus, the evaluation of emotional value and the automatic search for familiarity of an encountered face are specific to realistic conditions and are not reflected by results from the conventional laboratory.

5. Conclusion

The translation of conventional laboratory conditions into a novel, more realistic setup in VR presents a first step towards the investigation of real-life face perception. To our knowledge, this is the first study directly comparing face perception between a conventional 2D monitor setup and realistic conditions using VR.

In line with previous laboratory studies, our ERP analysis confirms that the N170 does seem to be face-specific, however, only

to a certain degree as it loses considerable discriminatory power in VR. These results raise doubt to the N170 as a meaningful marker for real-life face processing and our study implies it to be domain-specific, i.e., specific to the monitor. Specifically, its discriminatory capability is only applicable to planar, two-dimensional and unresponsive, but not real-life faces.

Our results on later components reveal distinct mechanisms for faces, silhouettes and objects being applied under VR conditions as opposed PC conditions. This is further supported by an in-depth source analysis suggesting a tripartite processing structure: First, early detection of perceptual face characteristics, second, registration of emotional value, and finally self-relevant retrieval of, and comparison with familiar faces of which any processing beyond basic perceptual properties is manifested in later components.

Hence, our study is in line with previous studies contrasting electrophysiological markers obtained under 2D with VR conditions, providing evidence that said markers and the functional neural properties they reflect are specific to the domain in which they occur. In a study on frontal-alpha-asymmetry (FAA) by [Schöne et al. \(2021a,b\)](#) it became apparent that the FAA does not index the same emotional and motivational state in both modalities, 2D and VR. Likewise, in a memory paradigm, the theta old-new-effect could not be replicated when the stimuli were first presented in VR ([Kisker et al., 2021b](#)). On a more general note, researchers should be aware that their findings do not necessarily translate to realistic conditions and should therefore be careful when generalizing their results beyond the setting they were observed ([Yarkoni, 2022](#)). Specifically, cognitive and emotional processes might not generalize beyond the conditions under which they are measured (see also [Schöne, 2022](#)).

To further investigate why and how the N170's face specificity seems to not be consistent when obtained under realistic conditions in VR, more research in this direction is needed. It would be of great interest to investigate face processing in VR further by comparing the perception of faces to other object categories (e.g., cars) and to give participants the task to actively discriminate faces from controls (e.g., button press). Taken from our results, it seems to be promising to take a look beyond the timeframe of the N170 – potentially towards later components such as the L1 – to find a real-life neural marker for face processing.

In summary, face perception is a complex interplay of neural mechanisms occurring on a broader timeframe than roughly 200 ms post stimulus presentation. The present study confirms conventional laboratory results, which provide first evidence for the relevance of late ERP components in face processing, and further extends this assumption to realistic conditions. By means of correlation analysis and identification of neural generators, we showed that realistic face perception includes early face detection, in parts captured by the N170. Beyond basic-level processing, however, face perception seems to require emotion assessment as well as self-relevant retrieval of and comparison with familiar faces, only reflected in late components that are only captured under realistic conditions.

5.1. Limitations

Even though VR as a research tool offers the possibility to increase realism under laboratory conditions, the experimental design used in the present study is nevertheless modeled after

conventional laboratory setups. It is not our ambition to introduce a parallel research discipline employing VR, but to stay in line with previous research results on face perception, and to gradually bring the conventional laboratory closer to reality. The sequential presentation of numerous static, unresponsive stimuli in a block design that are appearing suddenly in front of the participant is still physically implausible and does not correspond to a real-life scenario. Most importantly, however, it allows for comparison of ERPs between modalities. The implementation of dynamic faces within a meaningful context and moreover, the possibility to respond to them can further increase the realism of the experimental setup. Thereby, an even closer approximation to real-life face processing will be achieved. The present results should be extended by frequency analysis in addition to ERP analysis, which is a promising cognitive manifestation involved in face perception and characterization ([Zion-Golumbic and Bentin, 2007](#)). Moreover, inquiring subjective measures of participants, e.g., arousal, valence and presence, would give further insight into the participant's perception of the VR modality.

It should be mentioned at this point that a great deal of studies on the N170 focus on the comparison of faces with inverted controls or other morphological modifications that are not easily translatable to VR, simply because they would appear extremely irritating to the participant. Consequently, the comparability to studies using these kinds of controls is limited. However, it could be considered to implement a similar design in VR in future studies to investigate whether the same effects are obtainable under realistic conditions and to consider later components here as well.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://osf.io/y8c6q/?view_only=0d2fd8d6bd1e4351afe0deb5e3f4d3a4.

Ethics statement

The studies involving human participants were reviewed and approved by the local ethic committee of Osnabrueck University, Germany. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any identifiable images or data included in this article.

Author contributions

All authors contributed to the study concept and design. JK and MJ developed the Unity VR environment, while BS provided the 2D stimulus presentation program. Testing and data collection was performed by MS and JK. Main data analyses, interpretation and visualization were performed by MS under supervision of BS and TG. Source analysis was performed and respective results were drafted by JK. Additional data analysis using robust statistics was performed by SS, respective parts of the manuscript were also drafted by SS. MS drafted the manuscript, MJ and

JK revised the manuscript. BS and TG provided critical revisions. All authors approved the final version of the manuscript for submission.

Funding

We acknowledge support by Deutsche Forschungsgemeinschaft (DFG) and Open Access Publishing Fund of Osnabrück University.

Acknowledgments

The authors thank Meret Tetau, Saskia Karcher and Maurice Scholz for assistance with data collection.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships.

References

- Aminoff, E. M., Kveraga, K., and Bar, M. (2013). The role of the parahippocampal cortex in cognition. *Trends Cogn. Sci.* 17, 379–390. doi: 10.1016/j.tics.2013.06.009
- Blascovich, J., Loomis, J., Beall, A. C., Swin, K. R., Crystal, L., Inquiry, S. P., et al. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychol. Inq.* 13, 103–124. doi: 10.1207/S15327965PLI1302_01
- Blau, V. C., Maurer, U., Tottenham, N., and McCandliss, B. D. (2007). The face-specific N170 component is modulated by emotional facial expression. *Behav. Brain Funct.* 3, 1–13. doi: 10.1186/1744-9081-3-7
- Boehm, S. G., Dering, B., and Thierry, G. (2011). Category-sensitivity in the N170 range: a question of topography and inversion, not one of amplitude. *Neuropsychologia* 49, 2082–2089. doi: 10.1016/j.neuropsychologia.2011.03.039
- Bohbot, V. D., Allen, J. B., Dagher, A., Dumoulin, S. O., Evans, A. C., Petrides, M., et al. (2015). Role of the parahippocampal cortex in memory for the configuration but not the identity of objects: converging evidence from patients with selective thermal lesions and fMRI. *Front. Hum. Neurosci.* 9, 1–17. doi: 10.3389/fnhum.2015.00431
- Bombardieri, D., Schmid, P. C., Schmid Mast, M., Birri, S., Mast, F. W., Lobmaier, J. S., et al. (2013). Emotion recognition: The role of featural and configural face information. *Quarterly Journal of Experimental Psychology* 66, 2426–2442.
- Bosch-Bayard, J., Valdés-Sosa, P., Virues-Alba, T., Aubert-Vázquez, E., Roy John, E., Harmony, T., et al. (2001). 3D statistical parametric mapping of EEG source spectra by means of variable resolution electromagnetic tomography (VARETA). *Clin. Electroencephalogr.* 32, 47–61. doi: 10.1177/155005940103200203
- Bubb, E. J., Kinnavane, L., and Aggleton, J. P. (2017). Hippocampal–diencephalic–cingulate networks for memory and emotion: an anatomical guide. *Brain Neurosci. Adv.* 1:239821281772344. doi: 10.1177/2398212817723443
- Bublitzky, F., Gerdes, A. B. M., White, A. J., Riemer, M., and Alpers, G. W. (2014). Social and emotional relevance in face processing: happy faces of future interaction partners enhance the late positive potential. *Front. Hum. Neurosci.* 8, 1–10. doi: 10.3389/fnhum.2014.00493
- Bukach, C. M., Gauthier, I., and Tarr, M. J. (2006). Beyond faces and modularity: the power of an expertise framework. *Trends Cogn. Sci.* 10, 159–166. doi: 10.1016/j.tics.2006.02.004
- Busch, N. A., Debener, S., Kranczioch, C., Engel, A. K., and Herrmann, C. S. (2004). Size matters: effects of stimulus size, duration and eccentricity on the visual gamma-band response. *Clin. Neurophysiol.* 115, 1810–1820. doi: 10.1016/j.clinph.2004.03.015
- Civile, C., Elchlepp, H., McLaren, R., Galang, C. M., Lavric, A., and McLaren, I. P. L. (2018). The effect of scrambling upright and inverted faces on the N170. *Q. J. Exp. Psychol.* 71, 2464–2476. doi: 10.1177/1747021817744455
- De Cesare, A., and Codispoti, M. (2006). When does size not matter? Effects of stimulus size on affective modulation. *Psychophysiology* 43, 207–215. doi: 10.1111/j.1469-8986.2006.00392.x
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009
- Delorme, A., Sejnowski, T., and Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage* 34, 1443–1449. doi: 10.1016/j.neuroimage.2006.11.004
- Dering, B., Martin, C. D., Moro, S., Pegna, A. J., and Thierry, G. (2011). Face-sensitive processes one hundred milliseconds after picture onset. *Front. Hum. Neurosci.* 5, 1–14. doi: 10.3389/fnhum.2011.00093
- Dering, B., Martin, C. D., and Thierry, G. (2009). Is the N170 peak of visual event-related brain potentials car-selective? *Neuro Report* 20, 902–906. doi: 10.1097/WNR.0b013e328327201d
- Di Russo, F., Martínez, A., Sereno, M. I., Pitzalis, S., and Hillyard, S. A. (2002). Cortical sources of the early components of the visual evoked potential. *Hum. Brain Mapp.* 15, 95–111. doi: 10.1002/hbm.10010
- Eimer, M. (2000). The face-specific N170 component reflects late stages in the structural encoding of faces. *Neuro Report* 11, 2319–2324. doi: 10.1097/00001756-200007140-00050
- Eimer, M. (2011). The face-sensitivity of the N170 component. *Front. Hum. Neurosci.* 5, 1–2. doi: 10.3389/fnhum.2011.00119
- Evans, A. C., Collins, D. L., and Mills, S. R. (1993). 3D statistical neuroanatomical models from 305 MR I volumes. *IEEE Nucl Sci Symp Med Imag* 108:187721878. doi: 10.1109/NSSMIC.1993.373602
- Gabana, D., Tokarchuk, L., Hannon, E., and Gunes, H. (2018). Effects of Valence and Arousal on Working Memory Performance in Virtual Reality Gaming. 2017 7th International Conference on Affective Computing and Intelligent Interaction, ACII 2017, 2018-Janua, 36–41.
- Ganis, G., Smith, D., and Schendan, H. E. (2012). The N170, not the P1, indexes the earliest time for categorical perception of faces, regardless of interstimulus variance. *Neuro Image* 62, 1563–1574. doi: 10.1016/j.neuroimage.2012.05.043
- Giovanello, K. S., Alexander, M., and Verfaellie, M. (2003). Differential impairment of person-specific knowledge in a patient with semantic dementia. *Neurocase* 9, 15–26. doi: 10.1076/neur.9.1.15.14369
- Gromer, D., Madeira, O., Gast, P., Nehfischer, M., Jost, M., Müller, M., et al. (2018). Height simulation in a virtual reality cave system: validity of fear responses and effects of an immersion manipulation. *Front. Hum. Neurosci.* 12, 1–10. doi: 10.3389/fnhum.2018.00372
- Gruber, T., Trujillo-Barreto, N. J., Giabbiconi, C. M., Valdés-Sosa, P. A., and Müller, M. M. (2006). Brain electrical tomography (BET) analysis of induced gamma band responses during a simple object recognition task. *Neuro Image* 29, 888–900. doi: 10.1016/j.neuroimage.2005.09.004
- Ham, A. O., Schupp, H. T., and Weike, A. I. (2003). *Motivational Organization of Emotions: Autonomic Changes, Cortical Responses, and Reflex Modulation*. eds. R. J. Davidson, K. R. Scherer and H. H. Goldsmith Oxford University Press, 187–211.
- Herbert, C., Sfaerlea, A., and Blumenthal, T. (2013). Your emotion or mine: labeling feelings alters emotional face perception—an ERP study on automatic and intentional affect labeling. *Front. Hum. Neurosci.* 7, 1–14. doi: 10.3389/fnhum.2013.00378
- Herrmann, M. J., Ehlis, A. C., Ellgring, H., and Fallgatter, A. J. (2005a). Early stages (P 100) of face perception in humans as measured with event-related potentials (ERPs). *J. Neural Transm.* 112, 1073–1081. doi: 10.1007/s00702-004-0250-8

that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1050892/full#supplementary-material>

- Herrmann, M. J., Ehls, A. C., Muehlberger, A., and Fallgatter, A. J. (2005b). Source localization of early stages of face processing. *Brain Topogr.* 18, 77–85. doi: 10.1007/s10548-005-0277-7
- Hertweck, S., Weber, D., Alwanni, H., Unruh, F., Fischbach, M., Latoschik, M. E., et al. (2019). Brain Activity in Virtual Reality: Assessing Signal Quality of High-Resolution EEG while Using Head-Mounted Displays. 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 970–971.
- Ip, C., Wang, H., and Fu, S. (2017). Relative expertise affects N170 during selective attention to superimposed face-character images. *Psychophysiology* 54, 955–968. doi: 10.1111/psyp.12862
- Itier, R. J., and Taylor, M. J. (2002). Inversion and contrast polarity reversal affect both encoding and recognition processes of unfamiliar faces: a repetition study using ERPs. *Neuro Image* 15, 353–372. doi: 10.1006/nimg.2001.0982
- Itier, R. J., and Taylor, M. J. (2004). Source analysis of the N170 to faces and objects. *Neuro Report* 15, 1261–1265. doi: 10.1097/01.wnr.0000127827.73576.d8
- Jemel, B., Schuller, A. M., Cheref-Khan, Y., Goffaux, V., Crommelinck, M., and Bruyer, R. (2003). Stepwise emergence of the face-sensitive N170 event-related potential component. *Neuro Report* 14, 2035–2039. doi: 10.1097/00001756-200311140-00006
- Johnsdorf, M., Kisker, J., Gruber, T., and Schöne, B. (2023). Comparing encoding mechanisms in realistic virtual reality and conventional 2D laboratory settings: event-related potentials in a repetition suppression paradigm. *Front. Psychol.* 14, 1–12. doi: 10.3389/fpsyg.2023.1051938
- Johnston, P., Molyneux, R., and Young, A. W. (2015). The N170 observed “in the wild”: robust event-related potentials to faces in cluttered dynamic visual scenes. *Soc. Cogn. Affect. Neurosci.* 10, 938–944. doi: 10.1093/scan/nsu136
- Josef Golubic, S., Susac, A., Grilj, V., Ranken, D., Huonker, R., Hauelsen, J., et al. (2011). Size matters: MEG empirical and simulation study on source localization of the earliest visual activity in the occipital cortex. *Med. Biol. Eng. Comput.* 49, 545–554. doi: 10.1007/s11517-011-0764-9
- Keltner, D., Ekman, P., Gonzaga, G. C., and Beer, J. (2003). “Facial expression of emotion” in *Handbook of Affective Sciences*, eds. R. J. Davidson, K. R. Scherer and H. H. Goldsmith (Oxford University Press), 415–432.
- Keltner, D., and Kring, A. M. (1998). Emotion, social function, and psychopathology. *Rev. Gen. Psychol.* 2, 320–342. doi: 10.1037/1089-2680.2.3.320
- Key, A. P., and Corbett, B. A. (2020). The unfulfilled promise of the N170 as a social biomarker. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 5, 342–353. doi: 10.1016/j.bpsc.2019.08.011
- Kirasirova, L. A., Zakharov, A. V., Morozova, M. V., Kaplan, A. Y., and Pyatin, V. P. (2021). ERP correlates of emotional face processing in virtual reality. *Opera Med. Physiol.* 8, 12–19. doi: 10.24412/2500-2295-2021-3-12-19
- Kisker, J., Gruber, T., and Schöne, B. (2021a). Behavioral realism and lifelike psychophysiological responses in virtual reality by the example of a height exposure. *Psychol. Res.* 85, 68–81. doi: 10.1007/s00426-019-01244-9
- Kisker, J., Gruber, T., and Schöne, B. (2021b). Virtual reality experiences promote autobiographical retrieval mechanisms: electrophysiological correlates of laboratory and virtual experiences. *Psychol. Res.* 85, 2485–2501. doi: 10.1007/s00426-020-01417-x
- Kisker, J., Lange, L., Flinkenflügel, K., Kaup, M., Labersweiler, N., Tetenborg, F., et al. (2021c). Authentic fear responses in virtual reality: a Mobile EEG study on affective, behavioral and electrophysiological correlates of fear. *Front. Virtual Real.* 2, 1–19. doi: 10.3389/frvir.2021.716318
- Kozlovskiy, S. A., Pyasik, M. M., Korotkova, A. V., Vartanov, A. V., Glozman, J. M., and Kiselnikov, A. A. (2014). Activation of left lingual gyrus related to working memory for schematic faces. *Int. J. Psychophysiol.* 94:241. doi: 10.1016/j.jpsycho.2014.08.928
- Kuefner, D., de Heering, A., Jacques, C., Palmero-Soler, E., and Rossion, B. (2010). Early visually evoked electrophysiological responses over the human brain (P1, N170) show stable patterns of face-sensitivity from 4 years to adulthood. *Front. Hum. Neurosci.* 3, 67–72. doi: 10.3389/fpsyg.2010.0067.2009
- Latinus, M., and Taylor, M. J. (2006). Face processing stages: impact of difficulty and the separation of effects. *Brain Res.* 1123, 179–187. doi: 10.1016/j.brainres.2006.09.031
- Mair, P., and Wilcox, R. (2020). Robust statistical methods in R using the WRS2 package. *Behav. Res. Methods* 52, 464–488. doi: 10.3758/s13428-019-01246-w
- Martens, U., Trujillo-Barreto, N., and Gruber, T. (2011). Perceiving the tree in the woods: segregating brain responses to stimuli constituting natural scenes. *J. Neurosci.* 31, 17713–17718. doi: 10.1523/JNEUROSCI.4743-11.2011
- Matsumoto, D., Keltner, D., Shiota, M. N., O’Sullivan, M., and Frank, M. (2008). “Facial expressions of emotion” in *Handbook of Emotions*, eds. M. Lewis, J. M. Haviland-Jones and L. F. Barrett (The Guilford Press), 211–234.
- Mccarthy, G., Puce, A., Gore, J. C., and Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience* 605–610. doi: 10.1162/jocn.1997.9.5.605
- Miyashita, Y. (1993). INFERIOR TEMPORAL CORTEX: where visual perception meets memory. *Annu. Rev. Neurosci.* 16, 245–263. doi: 10.1146/annurev.ne.16.030193.001333
- Miyoshi, M., Katayama, J., and Morotomi, T. (2004). Face-specific N170 component is modulated by facial expressional change. *Neuro Report* 15, 911–914. doi: 10.1097/00001756-200404090-00035
- Mullen, T. R., Kothe, C. A. E., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., et al. (2015). Real-time neuroimaging and cognitive monitoring using wearable dry EEG. *I.E.E.E. Trans. Biomed. Eng.* 62, 2553–2567. doi: 10.1109/TBME.2015.2481482
- Nestor, A., Vettel, J. M., and Tarr, M. J. (2008). Task-specific codes for face recognition: how they shape the neural representation of features for detection and individuation. *PLoS One* 3:e3978. doi: 10.1371/journal.pone.0003978
- Newman, M., Gatersleben, B., Wyles, K. J., and Ratcliffe, E. (2022). The use of virtual reality in environment experiences and the importance of realism. *J. Environ. Psychol.* 79:101733. doi: 10.1016/j.jenvp.2021.101733
- Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Döbrowolny, H., and Panksepp, J. (2006). Self-referential processing in our brain – a meta-analysis of imaging studies on the self. *Neuro Image* 31, 440–457. doi: 10.1016/j.neuroimage.2005.12.002
- Ojeda, A., Bigdely-Shamlo, N., and Makeig, S. (2014). MoBILAB: an open source toolbox for analysis and visualization of mobile brain/body imaging data. *Front. Hum. Neurosci.* 8:121. doi: 10.3389/fnhum.2014.00121
- Palejwala, A. H., Dadario, N. B., Young, I. M., O’Connor, K., Briggs, R. G., Conner, A. K., et al. (2021). Anatomy and White matter connections of the lingual gyrus and cuneus. *World Neurosurg.* 151, e426–e437. doi: 10.1016/j.wneu.2021.04.050
- Pan, X., and Hamilton, A. F. (2018). Why and how to use virtual reality to study human social interaction: the challenges of exploring a new research landscape. *Br. J. Psychol.* 109, 395–417. doi: 10.1111/bjop.12290
- Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Front. Hum. Neurosci.* 9, 1–19. doi: 10.3389/fnhum.2015.00660
- Pfabisan, D. M., Sailer, U., and Lamm, C. (2015). Size does matter! Perceptual stimulus properties affect event-related potentials during feedback processing. *Psychophysiology* 52, 1238–1247. doi: 10.1111/psyp.12458
- Pierce, K., Haist, F., Sedaghat, F., and Courchesne, E. (2004). The brain response to personally familiar faces in autism: findings of fusiform activity and beyond. *Brain* 127, 2703–2716. doi: 10.1093/brain/awh289
- Pion-Tonachini, L., Kreutz-Delgado, K., and Makeig, S. (2019). ICLABEL: an automated electroencephalographic independent component classifier, dataset, and website. *Neuro Image* 198, 181–197. doi: 10.1016/j.neuroimage.2019.05.026
- Pyles, J. A., Verstynen, T. D., Schneider, W., and Tarr, M. J. (2013). Explicating the face perception network with White matter connectivity. *PLoS One* 8, e61611–e61612. doi: 10.1371/journal.pone.0061611
- Ratcliff, R., Philastides, M. G., and Sajda, P. (2009). Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *Proc. Natl. Acad. Sci. U. S. A.* 106, 6539–6544. doi: 10.1073/pnas.0812589106
- Recio, G., Sommer, W., and Schacht, A. (2011). Electrophysiological correlates of perceiving and evaluating static and dynamic facial emotional expressions. *Brain Res.* 1376, 66–75. doi: 10.1016/j.brainres.2010.12.041
- Rolls, E. T. (2019). The cingulate cortex and limbic systems for emotion, action, and memory. *Brain Struct. Funct.* 224, 3001–3018. doi: 10.1007/s00429-019-01945-2
- Rossion, B. (2014). Understanding face perception by means of human electrophysiology. *Trends Cogn. Sci.* 18, 310–318. doi: 10.1016/j.tics.2014.02.013
- Rossion, B., Gauthier, I., Tarr, M. J., Pierenne, D., Debatisse, D., and Despland, P. A. (1999). The N170 occipito-temporal component is delayed to inverted faces but not to inverted objects: electrophysiological evidence of face-specific processes in the human brain. *Neuro Image* 9, 69–74.
- Rossion, B., and Jacques, C. (2008). Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *Neuro Image* 39, 1959–1979. doi: 10.1016/j.neuroimage.2007.10.011
- Rossion, B., and Jacques, C. (2011). “The N170: understanding the time-course of face perception in the human brain” in *The Oxford Handbook of ERP Components*, 115–142.
- Rousset, G. A., Husk, J. S., Bennett, P. J., and Sekuler, A. B. (2008). Time course and robustness of ERP object and face differences. *J. Vis.* 8, 3.1–3.18. doi: 10.1167/8.12.3
- Rousset, G. A., Macé, M. J. M., and Fabre-Thorpe, M. (2004). Animal and human faces in natural scenes: how specific to human faces is the N170 ERP component? *J. Vis.* 4, 13–21. doi: 10.1167/4.1.2
- Schwaninger, A., Lobmaier, J. S., Wallraven, C., and Collishaw, S. (2009). Two routes to face perception: evidence from psychophysics and computational modeling. *Cognitive Science* 33, 1413–1440.
- Schindler, S., Zell, E., Botsch, M., and Kissler, J. (2017). Differential effects of face-realism and emotion on event-related brain potentials and their implications for the uncanny valley theory. *Sci. Rep.* 7, 1–13. doi: 10.1038/srep45003
- Schöne, B., Kisker, J., Sylvester, R. S., Radtke, E. L., and Gruber, T. (2021a). Library for universal virtual reality experiments (luVRe): a standardized immersive 3D/360° picture and video database for VR based research. *Curr. Psychol.* doi: 10.1007/s12144-021-01841-1

- Schöne, B., Sylvester, R. S., Radtke, E. L., and Gruber, T. (2021b). Sustained inattention blindness in virtual reality and under conventional laboratory conditions. *Virtual Reality* 25, 209–216. doi: 10.1007/s10055-020-00450-w
- Schöne, B., Wessels, M., and Gruber, T. (2019). Experiences in virtual reality: a window to autobiographical memory. *Curr. Psychol.* 38, 715–719. doi: 10.1007/s12144-017-9648-y
- Schöne, B. (2022). Commentary: a review on the role of affective stimuli in event-related frontal alpha asymmetry. *Front. Comput. Sci.* 4:994071. doi: 10.3389/fcomp.2022.994071
- Schöne, B., Kisker, J., Lange, L., Gruber, T., Sylvester, S., and Osinsky, R. (2023). The reality of virtual reality. *Front. Psychol.* 14:1093014. doi: 10.3389/fpsyg.2023.1093014
- Simons, D. J., and Chabris, C. F. (1999). Gorillas in our midst: sustained inattention blindness for dynamic events. *Perception* 28, 1059–1074. doi: 10.1068/p281059
- Snow, J. C., and Culham, J. C. (2021). The treachery of images: how realism influences brain and behavior. *Trends Cogn. Sci.* 25, 506–519. doi: 10.1016/j.tics.2021.02.008
- Stekelenburg, J. J., and De Gelder, B. (2004). The neural correlates of perceiving human bodies: an ERP study on the body-inversion effect. *Neuro Report* 15, 777–780. doi: 10.1097/00001756-200404090-00007
- Stolz, C., Endres, D., and Mueller, E. M. (2019). Threat-conditioned contexts modulate the late positive potential to faces—a mobile EEG/virtual reality study. *Psychophysiology* 56, e13308–e13315. doi: 10.1111/psyp.13308
- Tauscher, J. P., Schottky, F. W., Grogork, S., Bittner, P. M., Mustafa, M., and Magnor, M. (2019). Immersive EEG: Evaluating Electroencephalography in Virtual Reality. 26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019- Proceedings, 1794–1800.
- Thierry, G., Martin, C. D., Downing, P., and Pegna, A. J. (2007). Controlling for interstimulus perceptual variance abolishes N170 face selectivity. *Nat. Neurosci.* 10, 505–511. doi: 10.1038/nn1864
- Thierry, G., Pegna, A. J., Dodds, C., Roberts, M., Basan, S., and Downing, P. (2006). An event-related potential component sensitive to images of the human body. *Neuro Image* 32, 871–879. doi: 10.1016/j.neuroimage.2006.03.060
- Vizioli, L., Foreman, K., Rousselet, G. A., and Caldara, R. (2010). Inverting faces elicits sensitivity to race on the N170 component: a cross-cultural study. *J. Vis.* 10, 15.1–15.23. doi: 10.1167/10.1.15
- Vuilleumier, P., and Pourtois, G. (2007). Distributed and interactive brain mechanisms during emotion face perception: evidence from functional neuroimaging. *Neuropsychologia* 45, 174–194. doi: 10.1016/j.neuropsychologia.2006.06.003
- Weiner, K. S., and Zilles, K. (2016). The anatomical and functional specialization of the fusiform gyrus. *Neuropsychologia* 83, 48–62. doi: 10.1016/j.neuropsychologia.2015.06.033
- Williams, M. (2021). “Temporal Cortex,” in *The Neuropathology of Schizophrenia*. ed. M. Williams. doi: 10.1007/978-3-030-68308-5_5
- Xu, C., Demir-Kaymaz, Y., Hartmann, C., Menozzi, M., and Siegrist, M. (2021). The comparability of consumers’ behavior in virtual reality and real life: a validation study of virtual reality based on a ranking task. *Food Qual. Prefer.* 87:104071. doi: 10.1016/j.foodqual.2020.104071
- Yarkoni, T. (2022). The generalizability crisis. *Behav. Brain Sci.* 45:e1. doi: 10.1017/S0140525X20001685
- Zion-Golumbic, E., and Bentin, S. (2007). Dissociated neural mechanisms for face detection and configural encoding: evidence from N170 and induced gamma-band oscillation effects. *Cereb. Cortex* 17, 1741–1749. doi: 10.1093/cercor/bhl100



OPEN ACCESS

EDITED BY

Alyssa A. Brewer,
University of California, Irvine, United States

REVIEWED BY

Denis Delisle Rodriguez,
Santos Dumont Institute (ISD), Brazil
Keigo Ushiyama,
The University of Electro-Communications,
Japan

*CORRESPONDENCE

Jing Wang
✉ wangpele@gmail.com
Gang Liu
✉ gangliu_@zzu.edu.cn

[†]These authors have contributed equally to this work and share first authorship

RECEIVED 23 October 2022

ACCEPTED 30 May 2023

PUBLISHED 20 June 2023

CITATION

Wang W, Shi B, Wang D, Wang J and
Liu G (2023) Enhanced lower-limb motor
imagery by kinesthetic illusion.
Front. Neurosci. 17:1077479.
doi: 10.3389/fnins.2023.1077479

COPYRIGHT

© 2023 Wang, Shi, Wang, Wang and Liu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Enhanced lower-limb motor imagery by kinesthetic illusion

Weizhen Wang^{1†}, Bin Shi^{1†}, Dong Wang^{1†}, Jing Wang^{1*†} and Gang Liu^{2*†}

¹Institute of Robotics and Intelligent Systems, School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, China, ²Henan Key Laboratory of Brain Science and Brain-Computer Interface Technology, School of Electrical Engineering, Zhengzhou University, Zhengzhou, China

Brain-computer interface (BCI) based on lower-limb motor imagery (LMI) enables hemiplegic patients to stand and walk independently. However, LMI ability is usually poor for BCI-illiterate (e.g., some stroke patients), limiting BCI performance. This study proposed a novel LMI-BCI paradigm with kinesthetic illusion (KI) induced by vibratory stimulation on Achilles tendon to enhance LMI ability. Sixteen healthy subjects were recruited to carry out two research contents: (1) To verify the feasibility of induced KI by vibrating Achilles tendon and analyze the EEG features produced by KI, research 1 compared the subjective feeling and brain activity of participants during rest task with and without vibratory stimulation (V-rest, rest). (2) Research 2 compared the LMI-BCI performance with and without KI (KI-LMI, no-LMI) to explore whether KI enhances LMI ability. The analysis methods of both experiments included classification accuracy (V-rest vs. rest, no-LMI vs. rest, KI-LMI vs. rest, KI-LMI vs. V-rest), time-domain features, oral questionnaire, statistic analysis and brain functional connectivity analysis. Research 1 verified that induced KI by vibrating Achilles tendon might be feasible, and provided a theoretical basis for applying KI to LMI-BCI paradigm, evidenced by oral questionnaire (Q1) and the independent effect of vibratory stimulation during rest task. The results of research 2 that KI enhanced mesial cortex activation and induced more intensive EEG features, evidenced by ERD power, topographical distribution, oral questionnaire (Q2 and Q3), and brain functional connectivity map. Additionally, the KI increased the offline accuracy of no-LMI/rest task by 6.88 to 82.19% ($p < 0.001$). The simulated online accuracy was also improved for most subjects (average accuracy for all subjects: 77.23% > 75.31%, and average F1 score for all subjects: 76.4% > 74.3%). The LMI-BCI paradigm of this study provides a novel approach to enhance LMI ability and accelerates the practical applications of the LMI-BCI system.

KEYWORDS

brain-computer interface, lower-limb, motor imagery, kinesthetic illusion, vibratory stimulation, electroencephalogram

1. Introduction

Stroke disables or kills several million people each year (Beal, 2010; Katan and Luft, 2018). The disability, especially the lower limb hemiplegia, highly impacts the lives of individuals (Rea et al., 2014; Katan and Luft, 2018). Rehabilitation therapy is vital for helping the survivor regain as much use of his/her lower limbs as possible. Traditional therapy acts on the distal physical level to indirectly influence the brain's neural system, such as physical therapy and occupational therapy (Belda-Lois et al., 2011). However, these indirect therapies usually have poor efficiency.

BCI directly detects and modulates brain activity (Abiri et al., 2019; Saha et al., 2021). There are two main BCI strategies to improve the lives of individuals among stroke patients, i.e., assistive BCI and rehabilitative BCI (Mane et al., 2020). In the last decade, rehabilitative BCI has emerged as one of the promising tools for lower-limb motor function restoration by adjusting neuronal plasticity in affected neural circuits (Mane et al., 2020; Romero-Laiseca et al., 2020; Bobrova et al., 2021). In the field of BCI, a lower limb exoskeleton control system based on steady state visual evoked potentials (SSVEP) is an efficient BCI system, such as achieved accuracies of $91.3 \pm 5.73\%$ and an information transfer rate (ITR) of 32.9 ± 9.13 bits/min (Kwak et al., 2015). However, recent study reviewed that the flickers used to encode the BCI command must be sufficiently intense to obtain a high-quality SSVEP, which will engage a relatively large portion of the visual resources. Moreover, such irritating visual stimuli are not only irrelevant to users' subjective intent, but even disrupt users and make them feel uncomfortable. Therefore, even though this BCI system can work well to assist subjects with walking, its unnatural way of interaction is unacceptable to some users, which reduces its usefulness in practice (Xu et al., 2021).

LMI-BCI is a rehabilitative BCI and is natural. It detects the decrease/increase of power in the sensorimotor cortex to control external equipment and then reversely modulates brain activity by external equipment (Pfurtscheller and da Silva, 1999; Abiri et al., 2019; Zhang et al., 2021). Two types of MI can be distinguished: Kinesthetic Motor Imagery (KMI) and Visual Motor Imagery (VMI). A KMI can be described as the ability to imagine performing a movement without executing it, by imagining haptic sensations felt during the real movement (i.e., tactile, proprioceptive, and kinesthetic). In comparison, VMI represents a visualization of the corresponding movement incorporating the visual network (Rimbert et al., 2018; Yang et al., 2021). Nevertheless, there are three challenges for MI-BCI system: (1) EEG features is unstable in general; (2) It is difficult to detect lower-limb EEG features because the anatomical location of the lower-limb motor cortical area deep within the contralateral mesial cortex (Neuper and Pfurtscheller, 1996; Pfurtscheller et al., 1997; Pfurtscheller and da Silva, 1999; Boord et al., 2010); (3) Brain injury often inhibits the LMI ability after stroke, making the detection of EEG features harder (Takeda et al., 2007; Meyer et al., 2016; Park et al., 2016). To address the above challenges, several studies focused on enhancing LMI ability by Mirror Neuron system (MNS). MNS can transform visual sensory input of related behavior (e.g., Action Observation, AO) into one's own brain impulse or motor output of behavior (Kohler et al., 2002; Rizzolatti and Sinigaglia, 2016; Tanaka, 2021). Li et al. (2015) enhanced LMI ability in an imagining playing football task by Action Observation and demonstrated more distinctive features in LMI with AO than without. As a kind of action observation, virtual reality (VR) plays an essential role in enhancing neural activity during motor imagery. These studies suggest that the use of immersive virtual reality headsets, with the illusion and embodiment they provide, can effectively improve motor imagery training and BCI performance (Choi et al., 2020; Ferrero et al., 2021). In the field of lower-limb rehabilitation, standing and sitting are two regular movements. The experimental results of Triana-Guzman et al. indicated that the classification of motor imagery and idle state provided a mean accuracy of $88.51 \pm 1.43\%$ and $85.29 \pm 1.83\%$ for the sit-to-stand and stand-to-sit transitions, respectively (Triana-Guzman et al., 2022). Additionally, Chaisaen et al. (2020) classified the AO/MI

of sit-to-stand/stand-to-sit task, and the highest mean accuracy is $82.73 \pm 2.54\%$.

In 2021, a latest study revealed that mimicking known biological control principles results in BCI performance that is closer to healthy human abilities (Flesher et al., 2021). Brain-muscle-kinesthesia loop is the known biological control principle (Daly and Wolpaw, 2008). Hemiplegia inhibits kinesthesia after stroke (De Vries and Mulder, 2007; Mane et al., 2020). Kinesthetic illusion will complement the relevant loop. Neuroimaging studies have revealed that a relationship exists between the movement that has been imagined and the activation patterns of somatotopically organized motor and kinesthetic areas (Giannopulu and Mizutani, 2021). In the upper-limb MI-BCI study, KI induced by rubber hand can significantly amplify EEG features and provide better guidance to enhance upper-limb MI (Song and Kim, 2019). No previous study has investigated enhancing LMI-BCI performance by KI, although KI has important effects on BCI. This paper proposed an enhanced LMI paradigm by KI. Specifically, earlier studies proved that kinesthetic illusion where one feels muscle stretch could be induced by artificially vibrating the muscle spindle and tendon (Goodwin et al., 1972; Naito and Ehrsson, 2001; Naito, 2004; Tapin et al., 2021). Therefore, we induced KI by vibratory stimulation on Achilles tendon to enhance LMI ability during imagining kicking a football. We designed both research contents. The first research verified the feasibility of induced KI by vibratory stimulation and compared the brain activity during V-rest and rest task. The second research explored the effect of KI on LMI by comparing the LMI with and without KI (KI-LMI vs. no-LMI).

The rest of this paper is organized as follows. The experimental description and analysis methods are introduced in Section 2. Then, results are illustrated in Section 3 in terms of EEG analysis and BCI performance. Finally, discussion and conclusion are presented in Sections 4 and 5, respectively.

2. Materials and methods

2.1. Participants

Sixteen healthy subjects without a history of any neurological diseases (Subject H1-H16, all right-handed, age: 25 ± 2.98 years) participated in this study. Participants were asked to sleep normally and refrain from alcohol, caffeine and stimulant foods for 24 h before the experiment. All subjects were informed about the experimental process and required to sign an approved informed consent form before participating. This study was approved by the Ethics Committee of Xi'an Jiaotong University (Approval No. 2021-1,577).

2.2. Required equipment

Previous researches have revealed that activated cortex area (i.e., contralateral primary motor cortex) by KI was similar to the activated cortex area by MI (Naito, 2004; Lopez et al., 2017; Giannopulu and Mizutani, 2021). Therefore, EEG signal was recorded using 16-channel active electrodes placed over the sensorimotor cortex (Detailed locations are illustrated in Figure 1E), with the g.USBamp (g.tec Inc., Austria)

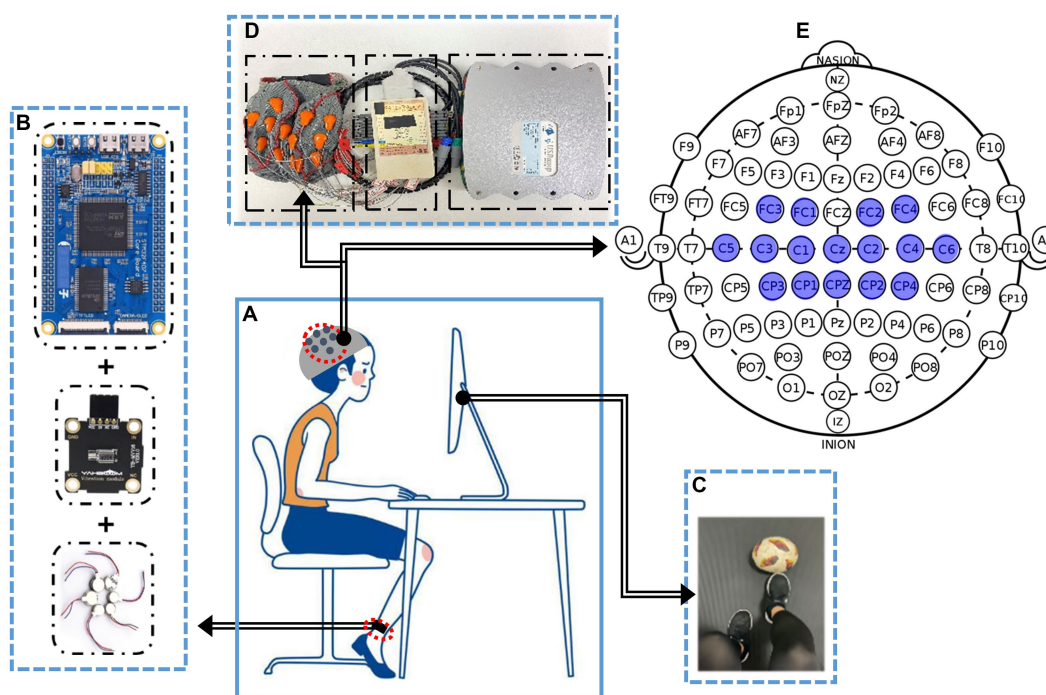


FIGURE 1

The experimental framework of enhanced LMI by KI. (A) Experiment condition, i.e., subjects were seated in a comfortable chair with feet resting on the floor in front of the monitor. (B) A inducing Illusion device. Kinesthetic illusion was induced by artificially vibrating the Achilles tendon. (C) Visually induced interactive interface. (D) EEG acquisition equipment. (E) Detailed channel locations are placed in the sensorimotor cortex (i.e., blue marked area).

system (See Figure 1D for the composition structure) according to the 10–10 electrode location system (Jurcak et al., 2007). The reference electrode and the ground electrode were placed on A1 and A2, with the sampling rate of 1,200 Hz. To reduce artifacts and power line interference, impedances for all electrodes were kept below 5 K Ω . Meanwhile, an online band-pass filter between 0 and 60 Hz and a notch filter between 48 and 52 Hz were applied on the raw EEG. During the task, subjects were sat in a comfortable chair with feet resting on the floor in front of the monitor as shown in Figure 1A.

In terms of visual induction, the Psychtoolbox-3 toolbox is utilized to design a visual induction interface (i.e., Figure 1C). In the aspect of inducing illusion device, the microcontroller (STM32F4) sends Pulse Width Modulation (PWM) regulation signals to a vibration regulation module, and then the eccentric vibrator (RswTech-motor-0827) generates specified frequency. Detailed structure is shown in Figure 1B. The vibration stimulation device in this study has a vibration frequency of 180 Hz under rated voltage. Additionally, Pacinian corpuscles in the mechanical receptor of human skin are sensitive to above 100 Hz frequency (Breitwieser et al., 2012). Therefore, a vibration frequency of 180 Hz was selected for experimental investigation in this study.

2.3. Data acquisition

Before the experiment, subjects were told to perform the LMI task of right limb. In addition, A inducing Illusion device was

placed at the right Achilles tendon to induce kinesthetic illusion in right limb (see Figure 1A). Before the EEG acquisition, each participant needs to undergo 5 min training session by motor execution to become familiar with the experimental task. In this study, the pre-training method was used to make the subjects better actively to complete the LMI task. During the data acquisition, all subjects were required to avoid actual movements for collecting high-quality EEG data. All trials for participants were completed in 1 day to reduce the EEG variability in different time periods. This study collected EEG data under four different conditions. Detailed paradigms flow is described in Figure 2: At the beginning ($t = -3$ - -1 s) of each trial, a white cross was displayed on the center of the screen to remind subjects to stay focused. Subsequently, a text cue (rest, V-rest, no-LMI, or KI-LMI) appeared for 1 s. When 'V-rest or rest' is observed, only the black background was shown to the subject who was executing rest task with or without vibratory stimulation synchronously (i.e., $t = 0$ –3.5 s). When 'KI- LMI or no-LMI' is observed, the designed visual guidance of kicking a football was shown to the subject who was executing LMI task with or without vibratory stimulation synchronously (i.e., $t = 0$ –3.5 s). There was a 4 s relaxation at the end of each trial. According to the random text cue, each subject randomly implemented the rest/LMI task (see Figure 2) in order to exclude the effect of task order on the results. In order to reduce the fatigue of the subjects, this study take a 3 min break after each run. Each subject performed a total of 12 runs, and each run consisted of 8 trials for rest task and 8 trials for LMI task. Thus, a total of 192 trials (i.e., $8 \times 2 \times 12 = 192$) were performed by each subject.

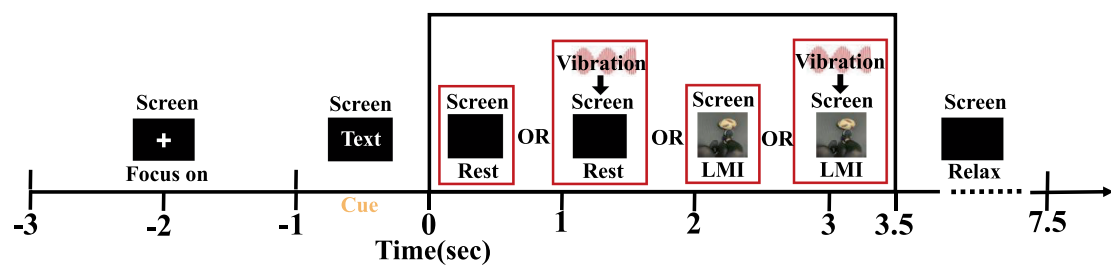


FIGURE 2

Overview of the different task paradigms. The process included four different tasks, i.e., rest, V-rest, noLMI or KI-LMI.

2.4. Research contents architecture

These collected data was applied to explore two research contents:

Research 1: This research aims to verify the feasibility of induced KI by vibrating Achilles tendon and analyze the EEG features produced by KI. Its process is as follows: (1) Oral questionnaire (Question1: Do you feel your feet moving during V-rest?) (2) The topographical distribution of rest and V-rest were compared to explore the independent effect of KI on cerebral cortex. (3) The EEG data of rest and V-rest were classified to quantify the independent effect of KI on rest task.

Research 2: This research aims to explore whether KI enhances LMI. Its process is as follows: (1) Topographical distribution, brain functional connectivity map and the ERD of Cz electrode (Cz-ERD) were analyzed during no-LMI and KI-LMI task. (2) The EEG data of rest and no-LMI/KI-LMI task were classified to explore the effect of KI on enhancing LMI. (3) Oral questionnaire (Question 2: Is KI conducive to focusing on LMI? Question 3: Which LMI do you prefer?) (4) Statistical test of C3/C4-ERD power explored whether KI conducive to distinguishing between left and right LMI.

2.5. Data analysis methods

2.5.1. Event related desynchronization analysis

This study selected 1 s EEG data as the baseline from the cross period (i.e., $t_{\text{base}} = -2.5 \sim -1.5$ s, see Figure 2), since the subjects were in the most relaxed state (i.e., the signal was the most stable) during t_{base} . Additionally, it selected 3 s EEG data as the task segment from the task period (i.e., $t_{\text{task}} = 0.5 \sim 3.5$ s), since severe artifacts were found within $0 \sim 0.5$ s. And, EEG data of the task period was used for preprocessing, feature extraction, and task classification. In order to enhance the quality of EEG data, all trials were visually inspected to remove data with more artifacts. The following exclusion criteria were applied to identify and discard noisy trials: (1) Maximum peak-to-peak value greater than $200 \mu\text{V}$; (2) the subject is blinking, the electrodes are not making good contact with the scalp, or there are some muscle artifacts (Delijorge et al., 2020; Triana-Guzman et al., 2022). Finally, any epoch where at least one electrode met these criteria was visually inspected to rule out noise-contaminated trials and labeled as an “artifact” manually.

Fourty trails of each task were randomly selected from the remaining data for each subjects. Event Related Desynchronization

(ERD) is caused the decrease of EEG frequency power in alpha (8–13 Hz) and beta (14–30 Hz) bands on the motor area related to the body parts by preparing movement. In contrast, Event Related Synchronization (ERS) is caused the increase of EEG frequency power in a similar way and bands (Machida and Tanaka, 2018). ERD power is one of the most common time-domain analysis methods for studying cerebral cortex activity during motor imagery. Therefore, we calculated the instantaneous power within a 0.25 s moving time window to describe the ERD change in the time domain (Phon-Amnuaisuk, 2008; Nakayashiki et al., 2014). In previous study (Pfurtscheller and da Silva, 1999; Graimann et al., 2002; Hashimoto and Ushiba, 2013), the classical method to compute the time course of ERD includes the following steps: (1) bandpass filtering of all event-related trials; (2) squaring of the amplitude samples to obtain power samples; (3) averaging of power samples across all trials; (4) averaging over time samples to smooth the data and reduce the variability. Similarly, the method is illustrated by equations (1–4) and applied to our research.

$$P_{\text{base}} = \frac{1}{T_{\text{base}}} \sum_{t \in T_{\text{base}}} P_t \quad (1)$$

$$P_{\text{task}} = \frac{1}{T_{\text{task}}} \sum_{t \in T_{\text{task}}} P_t \quad (2)$$

$$\text{ERD}(t) = \frac{P_t - P_{\text{base}}}{P_{\text{base}}} \times 100 \quad (3)$$

$$\overline{\text{ERD}} = \frac{P_{\text{task}} - P_{\text{base}}}{P_{\text{base}}} \times 100 \quad (4)$$

Where, P_t represents the instantaneous EEG power; P_{base} and P_{task} represent the average power during the period of T_{base} and T_{task} , respectively; $\text{ERD}(t)$ represents instantaneous ERD; $\overline{\text{ERD}}$ represents the average ERD of the task segment. Optimal time period of motor imagery were different due to the variability among subjects or trials. Furthermore, the $\overline{\text{ERD}}$ of each channel was used to draw topographical distribution and execute statistical test.

2.5.2. Functional connectivity analysis

To compare the patterns of two LMI tasks, we analyzed brain connectivity using the imaginary part of coherence (iCOH) algorithm. This algorithm is insensitive to artefactual caused by volume conduction, because a signal is not time-lagged to itself and thus manages to identify the synchronizations of two signals that are time-lagged (Nolte et al., 2004; Pezoulas et al., 2018). It is defined as:

$$\text{iCOH} = \text{imagin}(R_{x,y}(f)) \quad (5)$$

where $R_{x,y}(f)$ is the coherence of signals x , y , at f . Coherence is defined as the absolute value of coherency. The latter measures the linear relationship of the two signals at f . In fact, coherence acts as a generalization of correlation to the frequency domain with its values varying on the interval $[0, 1]$, where 1 indicates a perfect linear prediction of y from x .

In this study, the region of interest (ROIs) consists of 16 channels (Detailed channel locations are shown in Figure 1E). And, our study used iCOH algorithm between the signals at the paired cortical ROIs, and plotted the brain functional connectivity map (see Figure 3).

2.5.3. Feature extraction and classification algorithm

Common Spatial Patterns (CSP) is the well-known feature extraction method for analyzing EEG signals. However, its classification result lies on a certain frequency range. In fact, the optimal frequency band of motor imagery is different among subjects (Li et al., 2015; Liu et al., 2019; Shu et al., 2019). In recent years, an improved feature extraction algorithm called Filter Bank Common Spatial Pattern (FBCSP) has been applied to solve this deficiency. This method includes three stages (see Figure 4): Firstly, the target data is divided into different frequency bands by a band-pass filter. Secondly, the CSP algorithm is used to extract features of all sub-band data automatically. Lastly, the optimal feature selection is performed based on mutual information theory (Ang et al., 2008, 2012). In this paper, we divided 7–32 Hz frequency band into six sub-bands (i.e., 7–12, 12–16, 16–20, 20–24, 24–28, and 28–32 Hz). Finally, eight optimal target feature vectors are selected to achieve a better classification performance and reduce the information redundancy.

Support Vector Machine (SVM) with Gaussian kernel is suitable for classing the 2-class small samples of LMI (Yuan and Wang, 2008; Somadder and Saha, 2021). The Python platform is utilized to build a

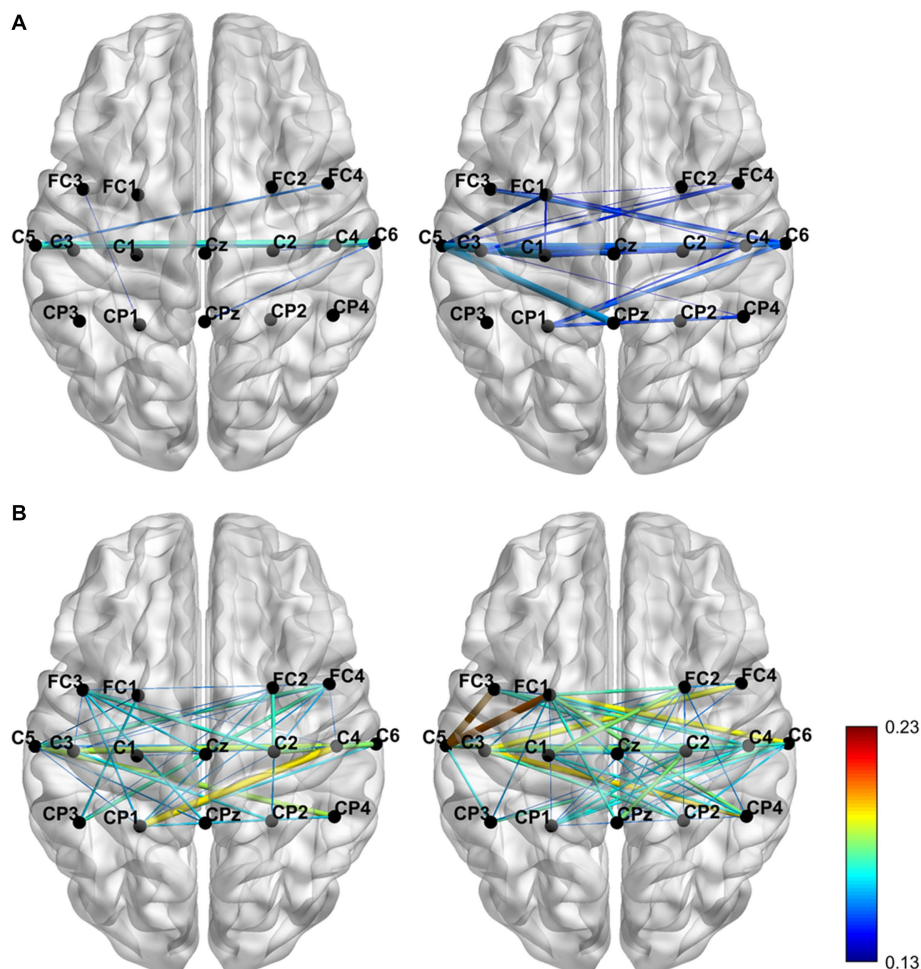


FIGURE 3

The brain functional connectivity map was grand averaged across subjects for the two conditions with the alpha (A) and beta (B) frequency bands separately.

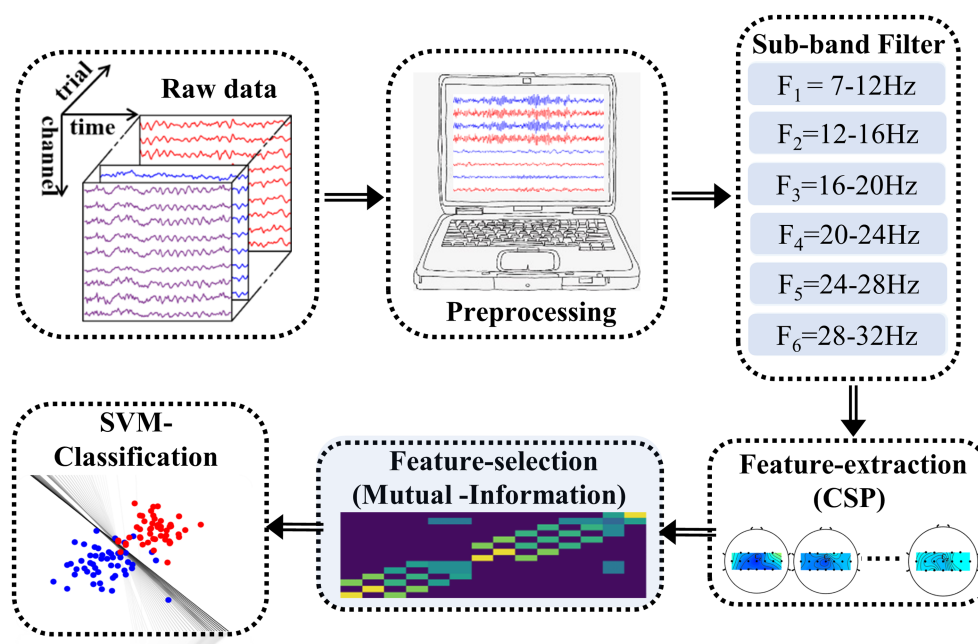


FIGURE 4

The Offline analysis process. This process includes the preprocessing, sub-band filter feature extraction (i.e., CSP), feature selection and classification.

TABLE 1 The confusion matrix of true-value and predicted-value.

Confusion matrix		True-value	
		1	0
Predicted-value	1	TP	FP
	0	FN	TN

suitable SVM model. The optimal parameters have a vital influence on classification performance. Thus, a cross-validation method is used to optimize the parameter C in the cost function and the parameter γ in the radial basis function.

2.6. Evaluation

This study used accuracy (Acc) and F1_score to evaluate classification performance and stability of the classification model. The process of how to calculate Acc and F1_score is as follows (Ren et al., 2020):

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (6)$$

$$\text{F1_score} = \frac{2 \cdot \frac{\text{TP}}{\text{TP} + \text{FP}} \cdot \frac{\text{TP}}{\text{TP} + \text{FN}}}{\frac{\text{TP}}{\text{TP} + \text{FP}} + \frac{\text{TP}}{\text{TP} + \text{FN}}} \quad (7)$$

Where, TP, TN, FN, and FP represent the relationship between the true-value and predicted-value (see Table 1).

In this study, we constructed four datasets (Dataset 1: Data of 40 rest trials and 40 V-rest trials; Dataset 2: Data of 40 rest trials and 40 no-LMI trials; Dataset 3: Data of 40 rest trials and 40 KI-LMI trials; Dataset 4: Data of 40 V-rest trials and 40 KI-LMI trials). Each data set is classified into two categories. Five-fold cross-validation is used to calculate the average accuracy.

The normality test results show that the significance level of the normality test is $p > 0.05$. EEG data fit a normality distribution. Therefore, statistical results of this paper were analyzed by paired T-test and One Sample T-test. These statistical test methods were calculated by using SPSS 24.0 mathematical tool. Subsequently, statistical graphs were drawn by MATLAB 2016 and GraphPad Prism 8.

3. Results

3.1. Research 1: verifying that KI is induced by vibrating Achilles tendon

This research aims to verify the feasibility of inducing KI via vibrating Achilles tendon and analyze the EEG features produced by KI.

3.1.1. Oral questionnaire (Q1)

After completing all the experimental tasks, each subject was given an oral questionnaire (Q1) to explore the correlation between vibration stimulation and KI. Its results are shown in Table 2. Firstly, we inquired Q1 that do you feel your feet moving slightly during V-rest. Thirteen subjects' answer is 'Yes'. Three subject's answer is 'No'. In practice, however, their feet were not moving. Therefore,

experiment 1 verified that KI (i.e., feel their feet moving) was felt subjectively by most subjects during vibrating Achilles tendon.

3.1.2. Topographical distribution of rest and V-rest

The first row of Figure 5 displays the grand-averaged topographical distribution for V-rest task relative to the rest task. In the period of V-rest task, there is a slight ERD phenomenon for Contralateral cerebral cortex (i.e., left hemispheric region) in both the β band and the $\alpha + \beta$ band, and the area around the Cz and CPz electrode (i.e.,

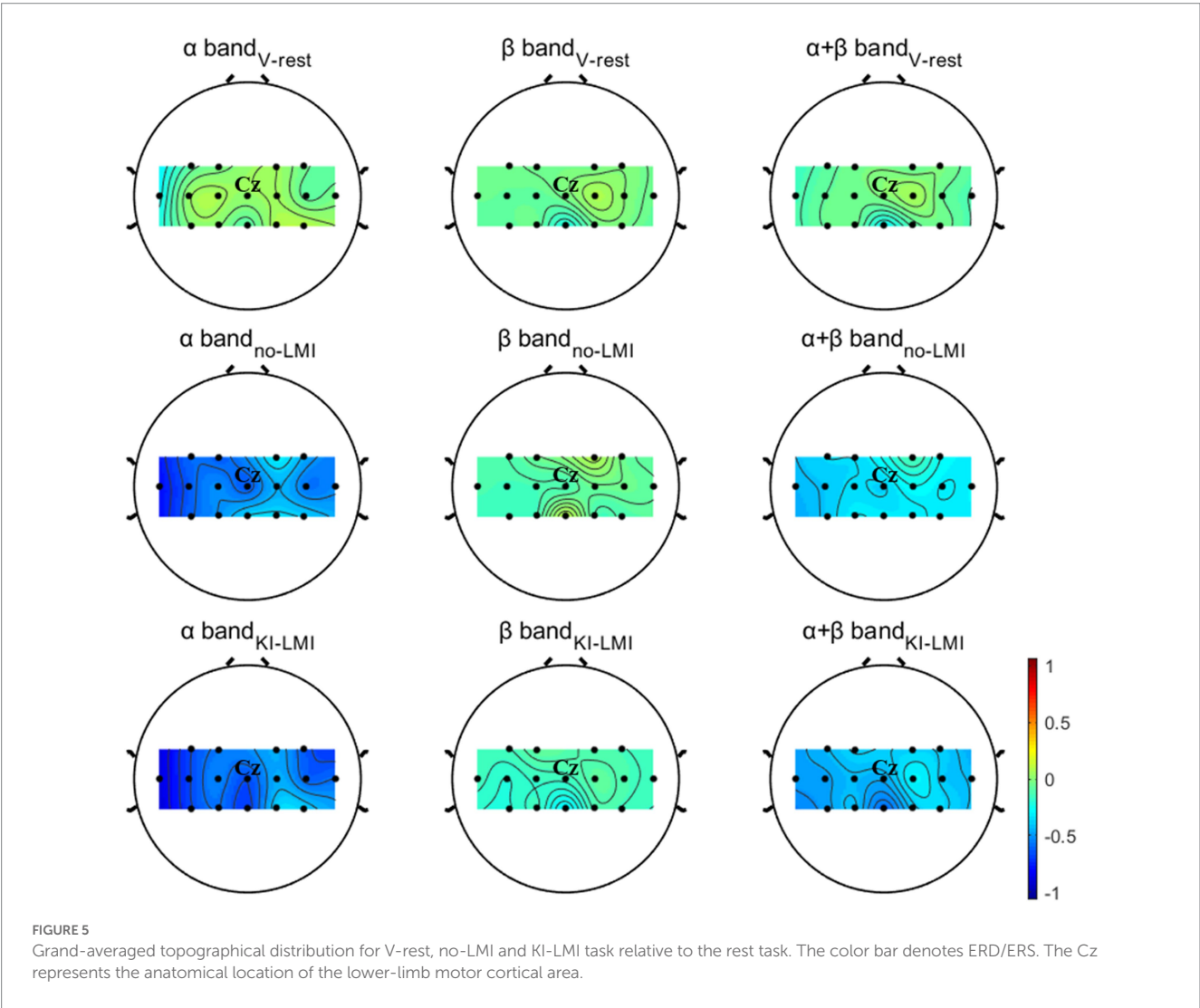
around the sensorimotor cortex) is predominantly activated. Therefore, induced KI by vibratory stimulation could not only be felt subjectively by subjects but also reflected in electrophysiological features.

3.1.3. Quantifying the independent effect of KI on rest task

The brain topographic maps of V-rest and rest show differences (see Figure 5). Our study classified V-rest and rest to quantify the independent effect of KI on rest task. Classification performance of

TABLE 2 The results of oral questionnaire.

Question	Content	Result (number)	
		No	Yes
Q1	Do you feel your feet moving slightly during V-rest?	No	Yes
		3	13
Q2	Is KI conducive to focusing on LMI task?	No	Yes
		4	12
Q3	Which LMI paradigm do you prefer?	no-LMI	KI-LMI
		4	12



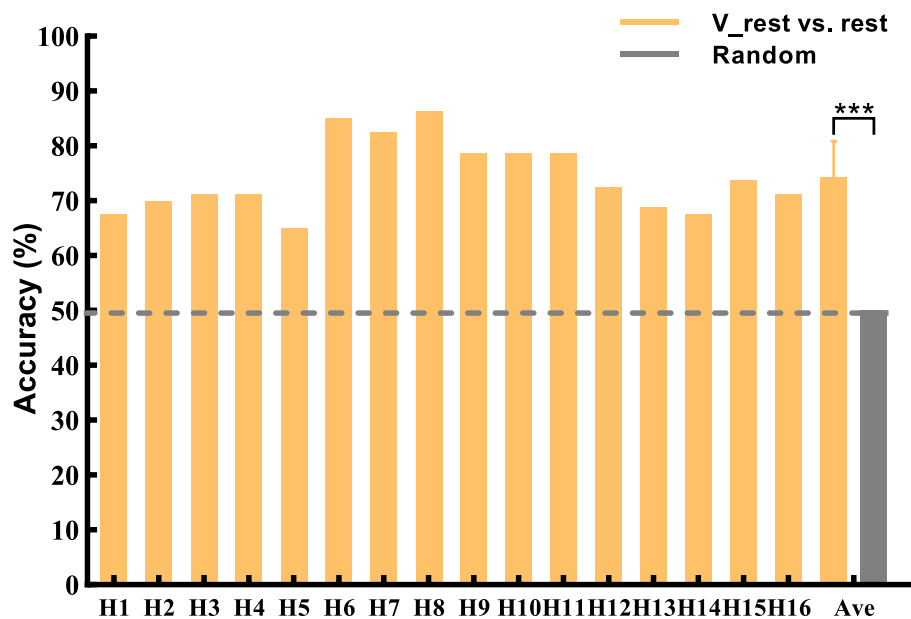


FIGURE 6

2-class accuracy of rest and V-rest. The black dashed line represents random (50%) of accuracy that means no difference between V-rest and rest. Data: mean \pm SD. *** $p < 0.001$.

rest and V-rest for all subjects is shown in Figure 6. Results were statistically evaluated using a one-sample T-test. The lowest and highest accuracy are 65.0 and 86.25%, respectively. The average accuracy of all subjects achieved 74.3%, significantly higher than the random accuracy (50%). Therefore, there were statistically significant EEG features which is generated by KI.

3.2. Research 2: exploring whether KI enhances LMI

3.2.1. Topographical distribution and Cz-ERD power of no-LMI and KI-LMI

The second and third rows of Figure 5 displays the grand-averaged topographical distribution for no-LMI and KI-LMI task relative to the rest task. In the period of no-LMI task, the significant activation region mainly occurred around the Cz electrode (i.e., around the sensorimotor cortex) in both the α band and the $\alpha + \beta$ band. Compared with the no-LMI task, the KI-LMI task can generate more obvious cortex activation at all frequency bands (i.e., significant ERD). And, there is most significant activation for the around Cz and CPz electrodes at all frequency bands. Especially, this activation is more localized to the contralateral side of the cerebral cortex during performing the KI-LMI task.

The Cz-ERD power of LMI under 8–30 Hz is analyzed (see Figure 7), since the above topographical distribution displays a stronger feature around the Cz electrode. The Cz-ERD power of no-LMI shows a significant decreasing trend at the beginning of the task except for Subject H3, H5, H7 and H13 (see Figure 7A). Especially, the ERD power of Subject H3, H7, and H13 shows a slight decreasing trend after adding KI to the LMI task (i.e., KI-LMI). For most subjects, the ERD tendency was more obvious and the lowest ERD power (i.e., E_{\min}) was lower during the KI-LMI task. Figure 7B displays the

average ERD of all subjects. These results manifest that the average ERD tendency is more obvious during the KI-LMI task. On the one hand, the E_{\min} value is significantly smaller during the KI-LMI task than the no-LMI task ($E_{\min KI} < E_{\min no}$, $p < 0.001$, see Figure 7D). On the other hand, T_{KI} is significantly shorter during the KI-LMI task than the no-LMI task ($T_{KI} < T_{no}$, $p < 0.01$, see Figure 7C).

Figure 8 shows the grand-averaged relative power of Cz electrode for V-rest, no-LMI and KI-LMI task relative to the rest task at α and β bands. The results in the α band included the following (see Figure 8A): (1) During the V-rest task, there was a slight decrease in relative power over the whole task period; (2) During the no-LMI task, there is a large relative power decline, and the peak value reaches -21 dB; (3) Compared with the no-LMI task, KI-LMI task produced a more significant relative power decline, with a peak value of -21.5 dB. The results in the β band included the following (see Figure 8B): (1) During the V-rest task, relative power decreased within 0 ~ 1 s and fluctuated around the baseline within 1 ~ 3 s; (2) Compared with the α band, the β band produced a similar power decline trend during the LMI, but had a smaller peak value (i.e., no-LMI: -1.7 dB, KI-LMI: -2.1 dB).

3.2.2. Results of brain functional connectivity analysis for no-LMI/KI-LMI task

The iCOH of functional connectivity patterns was analyzed to plot the functional connectivity map and explore the effect of KI on brain activity. The functional connectivity map (see Figure 3) was grand-averaged across subjects for the two conditions with the alpha and beta frequency bands separately. In the Figure 3, The color variation and thickness of the connecting lines indicate the strength of functional connectivity between channels. During the no-LMI task, the channels had weak functional connectivity in the α band and strong functional connectivity in the β band. During KI-LMI, KI enhances the functional connectivity between the channels in both α

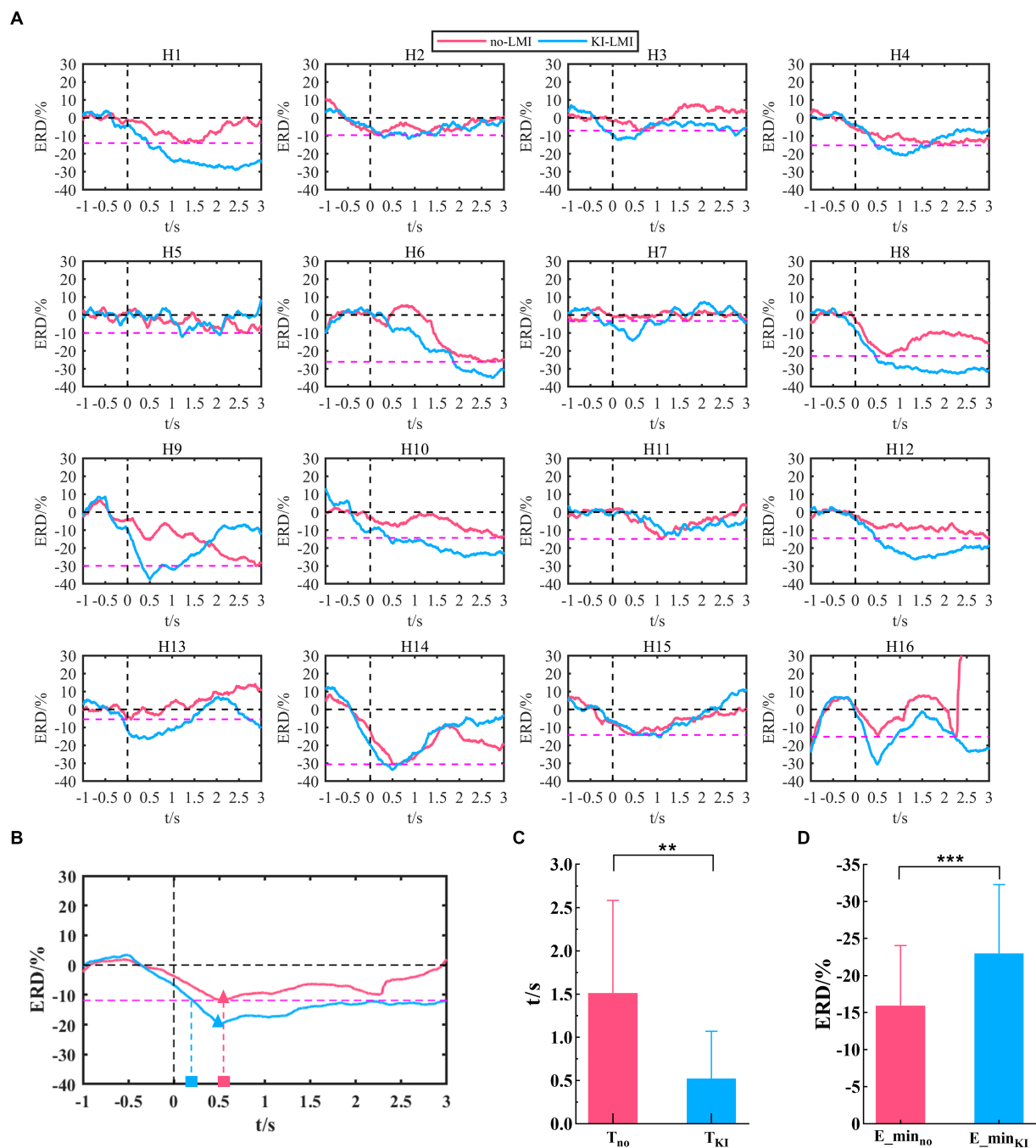


FIGURE 7

Cz-ERD power during LMI in the 8–30Hz band. (A) Cz-ERD power of the 16 subjects. The black vertical dashed line indicates the start of LMI. The black horizontal dashed line represents the baseline. The red dashed line represents the lowest ERD power ($E_{\min_{no}}$, $i=1, 2, 12$) during the no-LMI task for each subject i . (B) Average ERD of all subjects for no-LMI and KI-LMI tasks. The red/blue triangle represents the lowest ERD power ($E_{\min_{no}}$ or $E_{\min_{KI}}$) during the KI-LMI task. The red and blue rectangles indicate the time (T_{no} and T_{KI}) it takes to reach $E_{\min_{no}}$ for the LMI task. (C) and (D) Statistical test of $E_{\min_{no}}/E_{\min_{KI}}$ and T_{no}/T_{KI} . $E_{\min_{no}}$, $E_{\min_{KI}}$, T_{no} and T_{KI} represent the mean value of $E_{\min_{no}}$, $E_{\min_{KI}}$, T_{no} and T_{KI} for all subjects, respectively. Data: mean \pm SD. ** $p<0.01$, *** $p<0.001$.

and β bands. In particular, the contralateral cortical connectivity was enhanced, such as channels FC1, FC3, CP3, C1, C3, C5, Cz, etc. These results indicate that KI can improve the spatial feature distribution of the brain during performing LMI task, enhance the connectivity among cerebral cortical channels, and thus improve the information transmission process between brain regions.

3.2.3. Classification performance of rest/V-rest and no-LMI/KI-LMI

Figure 9 shows the comparison of classification accuracy across subjects for classifying LMI and rest task (i.e., no-LMI vs. rest and KI-LMI vs. rest). The subjects with less than 70% classification accuracy are defined as BCI-illiterate (Shu et al., 2017; Zhang et al., 2021). As

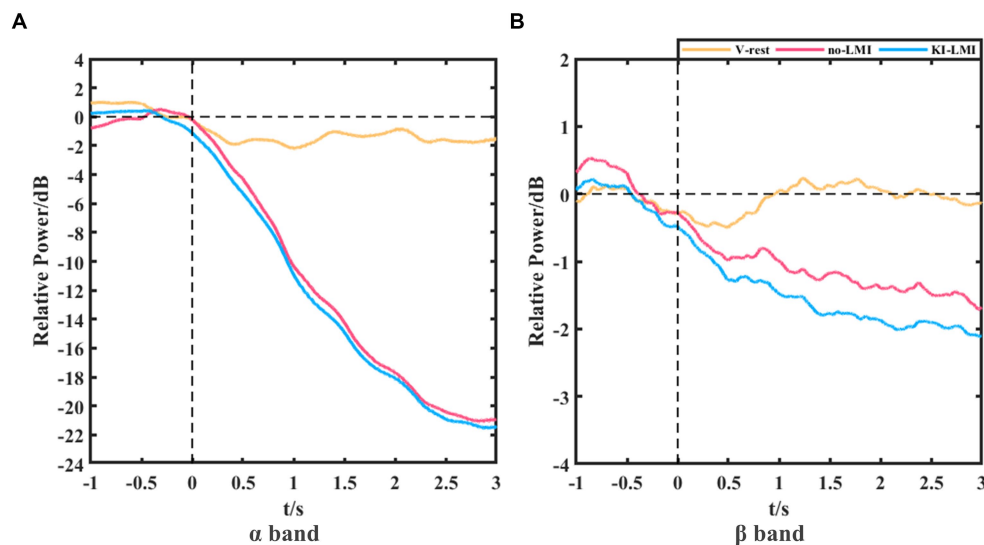


FIGURE 8

Grand-averaged relative power of Cz electrode for V-rest, no-LMI and KI-LMI task relative to the rest task at α band (A) and β band (B).

shown in Figure 9, six healthy subjects (H1, H2, H4, H11, H13 and H15) were BCI-illiterate during the no-LMI task. After adding the KI to the LMI (i.e., KI-LMI), average accuracy of these five subjects increased by 9.17%. In addition, KI generally improves LMI-BCI offline accuracy and F1_score except for subject H3, and the accuracy and F1_score of all subjects were greater than 70%. Especially, four subjects (H6, H8, H10 and H12) reached a higher BCI accuracy (> 85%). Classification results were statistically evaluated using a paired t-test method. The offline accuracy was significantly improved for classifying LMI and rest ($p < 0.001$), achieving a 6.88% improvement and reaching 82.19%. And, the results of Table 3 indicates that average F1_score value increased by 7.6% after adding KI to LMI task (i.e., KI-LMI task). Seeing Table 3 for detailed F1_score of each participant.

Additionally, Figure 9 and Table 3 shows the results of classifying V-rest and KI-LMI. After adding vibration to the LMI and rest task (i.e., KI-LMI and V-rest), the simulated online average classification accuracy and F1_score of most subjects (H1, H4, H6, H8, H9, H11, H12, H13, H14, H15 and H16) was improved (average accuracy: 79.84% > 74.09%, average F1_score: 79.24% > 72.6%), and there were significant differences in two groups ($p < 0.05$). And, the simulated online average performance for all subjects was presented in Figure 9 and Table 3 (i.e., average accuracy: 77.23 > 75.31%, and average F1_score: 76.4 > 74.3%).

3.2.4. Oral questionnaire (Q2 and Q3)

After completing all the experimental tasks, each subject was given an oral questionnaire (Q2 and Q3) to explore the superiority of this research paradigm. Its results are shown in Table 2. Twelve subjects thought that KI was more conducive to focusing on LMI task. Thus they preferred the KI-LMI paradigm. In contrast, four subjects thought that a slight vibration is not enough to produce KI to enhance their LMI ability. Additionally, one subjects (H3) thought that a strong vibration distracted his attention on LMI.

3.2.5. Statistical test of C3/Cz/C4-ERD power

The topographical distribution in Figure 5 shows more obvious contralateral cortex activation during right lower-limb MI task. In order to explore the activation effect of KI on different major channels, the paired T-test analyzed the ERD power of channels C3, Cz and C4 under the no-LMI/KI-LMI tasks (see Table 4). For the channels C3 and Cz, there are significant differences ($p = 0.006 < 0.01$ and $p = 0.007 < 0.01$) in α frequency band, and there are significant differences ($p = 0.001$ and $p = 0.000 < 0.001$) in $\alpha + \beta$ frequency band. There are significant differences ($p = 0.026 < 0.05$) in β frequency band at channel C3. Nevertheless, for the channel C4, all p -values are far bigger than 0.05, which indicates there is no significant difference.

4. Discussion

In the previous research, it was well known that brain injury often inhibits the LMI-ability after stroke, making the detection of EEG features harder (Takeda et al., 2007; Meyer et al., 2016; Park et al., 2016). The Higher BCI performance generally accelerates the patients' recovery process (Guan, 2016). Therefore, enhancing LMI ability of stroke patients is critical to achieve rehabilitation. The latest study revealed that mimicking known biological control principles could improve BCI performance for subjects (Fletcher et al., 2021). A natural way of interaction is acceptable to users, which increases its usefulness in practice (Xu et al., 2021). Previous studies used visual induction consistent with motor tasks to enhance LMI ability (Boord et al., 2010; Kitahara et al., 2017; Yu et al., 2018). Additionally, kinesthetic illusion (KI) is a type of proprioception that can complement the biological control loop. Two studies verified that a combination of upper-limb MI and KI feedback improved MI-BCI performance (Barsotti et al., 2018; Song and Kim, 2019). However, no previous study has investigated enhancing LMI ability by KI to design LMI-BCI paradigm. Therefore, based on the above researches, we designed two research contents to verify the feasibility of inducing KI via vibrating

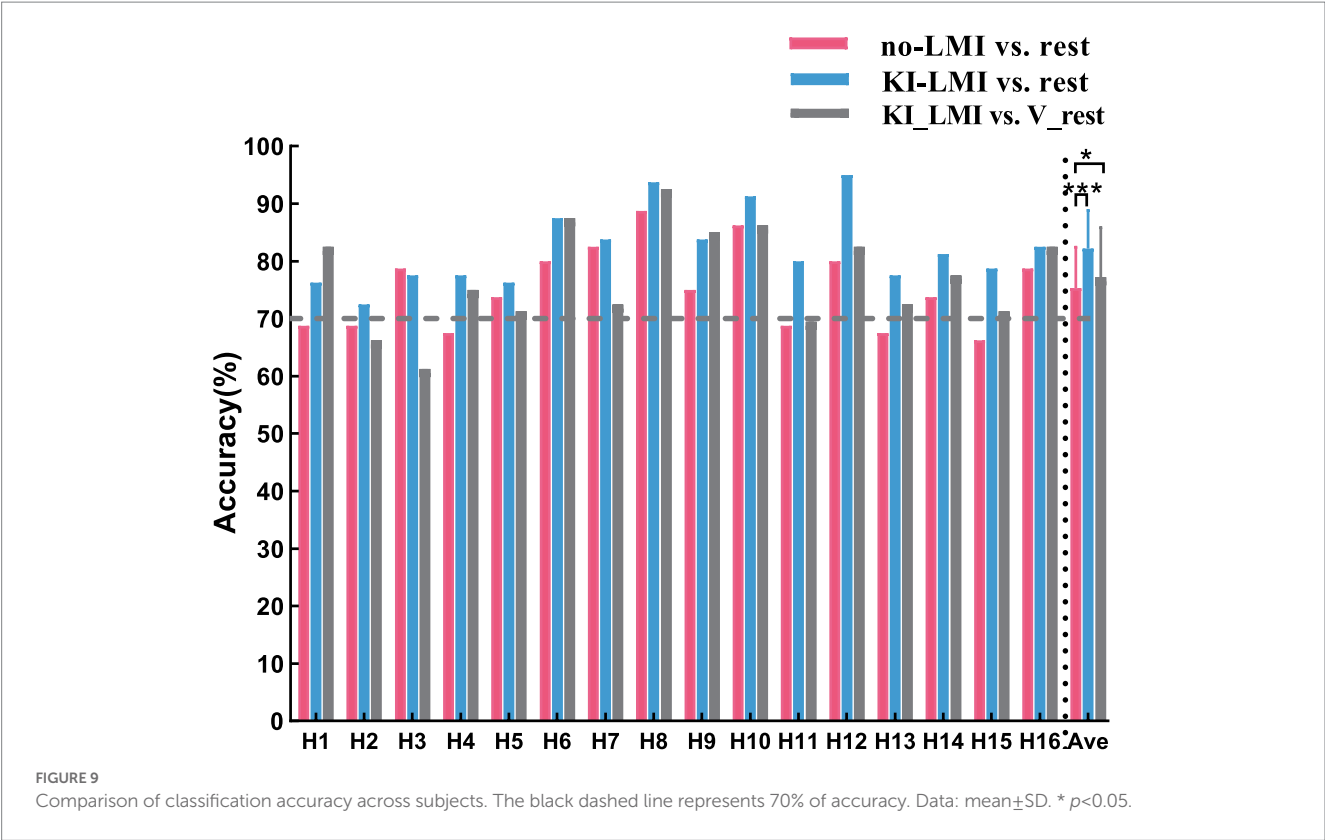


TABLE 3 Results of the F1_score across subjects.

Task	Subject/F1_score (%)																
	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14	H15	H16	Ave
Class1	63.9	66.8	78.7	65	75.0	80.5	83.0	87.7	70.8	86.7	65.1	80.2	66.9	74.1	64.9	79.5	74.3
Class2	76	70.1	74.9	76.1	77.3	87.9	84.2	93.8	82.1	92	80	94.8	77.3	81.7	78.5	83.7	81.9
Class3	81	65.5	58.3	71.1	69.1	88.2	72.3	92.3	83.9	86.1	69.3	81.4	72.9	79.1	68.9	83.5	76.4

Class1 represents the classification of rest and no-LMI; Class2 represents the classification of rest and KI-LMI; Class3 represents the classification of V-rest and KI-LMI; The bold number represents the maximum value for each subject.

Achilles tendon, analyze the EEG features produced by KI, and explore whether KI could enhance LMI ability.

Long-term research found that KI, where one feels muscle stretch, could be induced by artificially vibrating the muscle spindle and tendon of the limbs (Naito, 2004; Tapin et al., 2021). Neuroimaging studies have revealed that KI and MI have similar activation patterns in brain regions (Naito and Ehrsson, 2001; Naito, 2004; Giannopulu and Mizutani, 2021). In addition, vibration is better transmitted to the muscle spindle if applied over the tendon, where it results in longitudinal stretch of the muscle fibers (Taylor et al., 2017). Vibration stimulation of tendon is an effective way to induce KI (Barsotti et al., 2018; Tapin et al., 2021). In this study, the most obvious tendon location (i.e., Achilles tendon) was selected as the vibration stimulation point. In medical contexts, using a MI questionnaire as an ability predictor tool could be one possible way to estimate BCI performance (Vasilyev et al., 2017; Rimbart et al., 2018). In the field of BCI, two classical work studied two different MI questionnaires. The first study concludes that the KMI scores obtained from the Kinesthetic and Visual Imagery Questionnaire could predict the performance of a MI-based BCI for able-bodied subjects (Vuckovic and Osuagwu,

2013). The second study found that the representation of subjective behaviors, calculated using the Motor Imagery Questionnaire Revised-Second Edition, and the control of the BCI seem to be strongly linked (Marchesotti et al., 2016). These studies have shown that MI questionnaires are probably the most accepted and validated methods to measure the subjective feelings of a subject. Therefore, our study conducted an oral questionnaire for all subjects to explore the results of the experiment. In this study, firstly, an oral questionnaire (Q1) of Table 2 verified that KI was felt subjectively by most subjects (13 out of 16 subjects) during vibrating Achilles tendon. Secondly, KI induced by vibratory stimulation enhanced the activation of the area around the Cz and CPz electrodes (i.e., around the sensorimotor cortex) during the V-rest task (see Figure 5). This result is consistent with previous research that activated cortex area (i.e., contralateral primary motor cortex) by KI was similar to the activated cortex area by MI (Rosenkranz and Rothwell, 2003; De Moraes Silva et al., 2015; Lopez et al., 2017; Giannopulu and Mizutani, 2021; Zhang et al., 2021). Similarly, a recent study showed that vibrotactile neurofeedback training on upper limbs can increase motor cortical excitability in hand muscle representation corresponding to a muscle engaged by the

TABLE 4 Statistical results of the paired *t*-test.

Channel	P-value of no-LMI vs. KI-LMI		
	α (8–13Hz)	β (14–30Hz)	$\alpha+\beta$ (8–30Hz)
C3	0.006**	0.026*	0.001***
Cz	0.007**	0.094	0.000***
C4	0.866	0.148	0.191

This table compares the ERD power of two tasks in the same condition. Data: mean \pm SD.
 * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

MI (Grigorev et al., 2021). Lastly, the classification performance of rest and V-rest shows a significant difference (see Figure 6), which suggests that activated the independent effects on the cerebral cortex by KI can be classified clearly. This experimental study verified that induced KI *via* vibrating Achilles tendon may be feasible, and provided a theoretical basis for applying KI to LMI-BCI paradigm. In order to better explore the difference and correlation between LMI and KI, we need to ensure the same duration of LMI and KI, so this study chose the same induction duration as LMI (i.e., 3.5 s). The above results verified the effectiveness of KI induction with a duration of 3.5 s. However, there are still some defects in this study. 3.5 s may not be the best KI induction duration. Future studies will investigate the optimal duration of KI induction.

Both studies demonstrated that a combination of upper-limb MI and KI feedback improved MI-BCI performance (Barsotti et al., 2018; Song and Kim, 2019). In our study, we used many analytical methods (e.g., ERD power, topographical distribution, oral questionnaire, and functional connectivity analysis) to explore whether KI enhances LMI ability. Previous studies explored functional connectivity of different brain regions based on Granger causality analysis. These studies found that the significant causal connection from the visual area to the motor area under the “visual–auditory context” and the “visual context” may indicate the information transmission process of the dorsal pathway evoked by the visual stimulus. Therefore, our study used iCOH algorithm to analyze the brain functional connectivity between different channels. The results in Figure 3 indicate that KI can enhance the information transmission process between brain regions (especially the sensorimotor area) during performing the LMI task, thus improving LMI ability. In previous mechanism studies, the anatomical location of the lower limbs motor cortex deep within the central cortex of the interhemispheric fissure, which corresponds to the Cz electrode of the 10/10 standard electrode distribution system (Neuper and Pfurtscheller, 1996; Jurcak et al., 2007; Boord et al., 2010). Therefore, in our study about LMI, the Cz electrode was selected to analyze the ERD trend in the 8–30 Hz band during LMI tasks. When adding KI to the LMI task (i.e., KI-LMI), the cortex activation around the Cz electrode is more significant (see Figure 5), and the Cz-ERD tendency was more obvious for most subjects (see Figure 7A). To evaluate the MI enhancements of the proposed paradigm, we compared the observed ERD for different LMI tasks. The peak ERD amplitudes (i.e., $E_{\min}^1_{\text{no/KI}}$) were used to quantitatively evaluate the enhancement due to more immersive LMI from the paradigms (Song and Kim, 2019). Moreover, the ERD arrival time (i.e., $T^1_{\text{no/KI}}$) was adopted as an indicator of temporal characteristics of ERD because it was used to evaluate the

appropriateness of ERD for BCI, which is also an MI enhancement target for BCI system (Duann and Chiou, 2016; Song and Kim, 2019). In our study, T_{KI} is significantly shorter during the KI-LMI task than the no-LMI task ($T_{\text{KI}} < T_{\text{no}}$, see Figure 7C). Therefore, the ERD feature of KI-LMI was detected more quickly (i.e., better real-time; Duann and Chiou, 2016; Song and Kim, 2019), then the KI may speed up the online detection of LMI-BCI. The above significant ERD indexes are beneficial to EEG feature detection, thus improving the classification performance of BCI system. Referring to previous research methods (Storzer et al., 2016), this study plotted the grand-averaged relative power of Cz electrode (see Figure 8), in order to more clearly compare the effect of KI on different tasks. The results of Figure 8 and Figure 5 are consistent, and both show that relative power in the α band presents a more significant decline trend than that in the β band. This result indicated that the optimal frequency band of performing LMI is 8–13 Hz for some subjects in this study. And, above results suggest that KI enhances the relative power decline for all frequency bands during performing LMI and improves LMI ability of subjects.

In terms of BCI performance, KI significantly improved the average accuracy of classifying LMI and rest task (see Figure 9), achieving a 6.88% improvement. Especially, it is noteworthy that the average accuracy of all BCI-illiterate increased by 9.17%. Therefore, KI could enhance LMI, and enhanced LMI by KI may be more suitable for the BCI-illiterate (e.g., the partial stroke patients; Shu et al., 2017, 2019; Ren et al., 2020). Considering that if KI is applied to online LMI-BCI, vibration stimulation need to be performed simultaneously during LMI and rest tasks. Therefore, KI-LMI and V-rest are classified in our study to simulated online BCI performance. The results of this study show that KI is more conducive to the online classification of LMI and rest tasks for most subjects (see Figure 9, KI-LMI vs. V-rest, $p < 0.05$). Grand-averaged topographical distribution indicates that EEG features of KI-LMI and V-rest tasks have significant differences (i.e., More significant ERD feature were observed under the KI-LMI task). Therefore, better results can be obtained by classifying KI-LMI and V-rest tasks. Obviously, the high classification performance of KI-LMI/V-rest tasks is more suitable for the practical application of LMI-BCI paradigm. In addition, the results of Figure 7 and Figure 8 show that KI can enhance ERD feature and relative energy changes for most subjects. This result is more helpful for CSP to extract the two types of features with significant differences, so as to improve the classification performance (see Figure 9). In contrast, one subject (H3) thought that KI distracted his attention during the LMI task. Our research analyzed that higher vibration frequency might make this subject nervous because the Cz-ERD power of this subject was relatively unstable before the LMI task (see Figure 7A). The offline accuracy of the H3 is decreased by 1.25% (see Figure 9) due to the above reasons. Our research paradigm presented significant performance advantages for most subjects. Consequently, it may be more conducive to the training and rehabilitation of most stroke patients. Previous studies have found that activated areas of the cerebral cortex are biased toward the contralateral sensorimotor cortex during performing the LMI and visual observation task (Li et al., 2015; Yu et al., 2018). In this research (Tariq et al., 2020), LDA, SVM, and KNN model were used to classify bilateral foot LMI task. And, the single trial analysis and classification models resulted in high discrimination accuracies, i.e., maximum 83.4% for beta-ERS, 79.1% for beta-ERD, and 74.0% for mu-ERD. The above results are consistent

with our study in Figure 5. Additionally, our statistical results in Table 4 proved that the KI significantly enhanced the cortex activation at the left and middle region of sensorimotor cortex (i.e., channels C3 and Cz) during the LMI task. This result is consistent with the topographical distribution (see Figure 5). Therefore, the KI may be conducive to distinguishing the left/right LMI task. This study provides a new idea to design the left/right LMI-BCI paradigm.

Previous studies have shown that 70 to 80 Hz is an effective way to induce KI, and high vibration frequencies cannot induce KI (Naito et al., 1999; Naito, 2004; Taylor et al., 2017). The results of our study indicate that 180 Hz can induce lower-limb KI, which is somewhat different from previous results. Different vibration stimulation locations will induce different KI effects. Previous studies explored the optimal stimulation frequency range through vibration of the upper limb tendon (Naito et al., 1999; Naito, 2004; Taylor et al., 2017; Barsotti et al., 2018), while this study explored the effect of inducing lower-limb KI through vibration of Achilles tendon. The cerebral cortex region corresponding to the upper-limb and lower-limb has a great difference in physiological structure (i.e., the anatomical location of the lower-limb motor cortical area is smaller and deep within the contralateral mesial cortex; the lower limb is located at the distal physical level) (Michael, 2016), which makes the optimal vibration frequency for inducing the upper-limb and lower-limb KI may have great difference. The experimental results of our study verified the feasibility of lower-limb KI induced by vibration of Achilles tendon at 180 Hz, which provides some reference for the research on inducing lower-limb KI with vibration stimulation. However, the vibration parameters (e.g., vibration frequency and vibration duration) were not optimized in this study. In the future, the influence of different vibration parameters on inducing lower-limb KI will be explored to produce the optimal induction effect. Studies have shown that MI-BCI is beneficial to improve the rehabilitation of stroke patients (Mane et al., 2020). Therefore, our research paradigm further improves the rehabilitation effect and lower-limb motor function of stroke patients by improving the performance of LMI-BCI. However, our study removed data with more artifacts before classification, which may affect the results of online classification in LMI-BCI. Future research will design online experiments to explore the effect of removing data with more artifacts on online application. Although the obtained results are statistically significant, it is acknowledged that the sample size of the current study is limited to 16 healthy subjects. Therefore, more healthy subjects and stroke patients should be recruited to further study the features in sensorimotor cortex caused by KI in different groups.

5. Conclusion

In this study, we integrated the induced KI by vibrating Achilles tendon with LMI tasks to enhance LMI ability. Our research 1 has verified that the KI was induced by vibratory stimulation on Achilles tendon, and found that the activated cortex area by the KI was similar to the activated cortex area by LMI. Research 2 demonstrated that the KI could enhance the activation of the sensorimotor cortex and improve the classification performance of offline and simulated online LMI-BCI. Additionally, this study found that KI improved the

ipsilateral difference of cerebral cortex. The LMI-BCI paradigm of this study is conducive to enhance LMI ability and provides a new idea to design a novel left/right LMI-BCI paradigm. This proposed approach enriched the content of LMI-BCI technology and accelerated its research progress.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of Xi'an Jiaotong University (Approval No. 2021–1577). The participants provided their written informed consent to participate in this study.

Author contributions

WW, BS, and JW contributed to conception and design of this study. WW and DW carried out collecting EEG data and statistical analysis. BS, GL, and JW provided many suggestions for this study. WW and GL worked together to complete the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

Funding

This work was supported by the open project of Henan Key Laboratory of Brain Science and Brain-Computer Interface Technology under Grant HNBBL230101 and HNBBL230207, and the Shenzhen Science and Technology Plan projects under Grant JSGG20201102145602006, and 2021 Basic Research Project of Shenzhen Science, Technology and Innovation Commission under Grant JCYJ20210324123414039.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abiri, R., Borhani, S., Sellers, E. W., Jiang, Y., and Zhao, X. (2019). A comprehensive review of EEG-based brain-computer interface paradigms. *J. Neural Eng.* 16:011001. doi: 10.1088/1741-2552/aaf12e
- Ang, K. K., Chin, Z. Y., Wang, C., Guan, C., and Zhang, H. (2012). Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. *Front. Neurosci.* 6:39. doi: 10.3389/fnins.2012.00039
- Ang, K. K., Chin, Z. Y., Zhang, H., and Guan, C. (2008). *Filter Bank common spatial pattern (FBCSP) in brain-computer Interface*. International joint conference on neural networks, Hong Kong: People's Republic of China.
- Barsotti, M., Leonardi, D., Vanello, N., Bergamasco, M., and Frisoli, A. (2018). Effects of continuous Kinaesthetic feedback based on tendon vibration on motor imagery BCI performance. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 105–114. doi: 10.1109/TNSRE.2017.2739244
- Beal, C. C. (2010). Gender and stroke symptoms: a review of the current literature. *J. Neurosci. Nurs.* 42, 80–87. doi: 10.1097/JNN.0b013e3181ce5c70
- Belda-Lois, J.-M., Mena-del Horno, S., Bermejo-Bosch, I., Moreno, J. C., Pons, J. L., Farina, D., et al. (2011). Rehabilitation of gait after stroke: a review towards a top-down approach. *J. Neuroeng. Rehabil.* 8:66. doi: 10.1186/1743-0003-8-66
- Bobrova, E. V., Reshetnikova, V. V., Vershinina, E. A., Grishin, A. A., Bobrov, P. D., Frolov, A. A., et al. (2020). Use of imaginary lower limb movements to control brain-computer Interface systems. *Neurosci. Behav. Physiol.* 50, 585–592. doi: 10.1007/s11055-020-00940-z
- Boord, P., Craig, A., Tran, Y., and Nguyen, H. (2010). Discrimination of left and right leg motor imagery for brain-computer interfaces. *Med. Biol. Eng. Comput.* 48, 343–350. doi: 10.1007/s11517-010-0579-0
- Breitwieser, C., Kaiser, V., Neuper, C., and Müller-Putz, G. R. (2012). Stability and distribution of steady-state somatosensory evoked potentials elicited by vibro-tactile stimulation. *Med. Biol. Eng. Comput.* 50, 347–357. doi: 10.1007/s11517-012-0877-9
- Chaisaen, R., Autthasan, P., Mingchinda, N., Leelaarporn, P., Kunaseth, N., Tammajarung, S., et al. (2020). Decoding EEG rhythms during action observation, motor imagery, and execution for standing and sitting. *IEEE Sensors J.* 20, 13776–13786. doi: 10.1109/JSEN.2020.3005968
- Choi, J. W., Kim, B. H., Huh, S., and Jo, S. (2020). Observing actions through immersive virtual reality enhances motor imagery training. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 1614–1622. doi: 10.1109/TNSRE.2020.2998123
- Daly, J. J., and Wolpaw, J. R. (2008). Brain-computer interfaces in neurological rehabilitation. *Lancet Neurol.* 7, 1032–1043. doi: 10.1016/S1474-4422(08)70223-0
- De Moraes Silva, J., Lima, F. P. S., De Paula, A. R., Teixeira, S., Do Vale Bastos, V. H., Dos Santos, R. P. M., et al. (2015). Assessing vibratory stimulation-induced cortical activity during a motor task—a randomized clinical study. *Neurosci. Lett.* 608, 64–70. doi: 10.1016/j.neulet.2015.09.032
- De Vries, S., and Mulder, T. (2007). Motor imagery and stroke rehabilitation: a critical discussion. *J. Rehabil. Med.* 39, 5–13. doi: 10.2340/16501977-0020
- Deljorge, J., Mendoza-Montoya, O., Gordillo, J. L., Caraza, R., Martinez, H. R., and Antelis, J. M. (2020). Evaluation of a P300-based brain-machine Interface for a robotic hand-orthosis control. *Front. Neurosci.* 14:589659. doi: 10.3389/fnins.2020.589659
- Duann, J.-R., and Chiou, J.-C. (2016). A comparison of independent event-related desynchronization responses in motor-related brain areas to movement execution, movement imagery, and movement observation. *PLoS One* 11:e0162546. doi: 10.1371/journal.pone.0162546
- Ferrero, L., Ortiz, M., Quiles, V., Ianez, E., and Azorin, J. M. (2021). Improving motor imagery of gait on a brain-computer Interface by means of virtual reality: a case of study. *Ieee Access* 9, 49121–49130. doi: 10.1109/ACCESS.2021.3068929
- Flesher, S. N., Downey, J. E., Weiss, J. M., Hughes, C. L., Herrera, A. J., Tyler-Kabara, E. C., et al. (2021). A brain-computer interface that evokes tactile sensations improves robotic arm control. *Science* 372, 831–836. doi: 10.1126/science.abd0380
- Giannopulu, I., and Mizutani, H. (2021). Neural kinesthetic contribution to motor imagery of body parts: tongue, hands, and feet. *Front. Hum. Neurosci.* 15:602723. doi: 10.3389/fnhum.2021.602723
- Goodwin, G. M., McCloskey, D. I., and Matthews, P. B. C. (1972). Proprioceptive illusions induced by muscle vibration—contribution by muscle-spindles to perception. *Science* 175, 1382–1385. doi: 10.1126/science.175.4028.1382
- Graimann, B., Huggins, J. E., Levine, S. P., and Pfurtscheller, G. (2002). Visualization of significant ERD/ERS patterns in multichannel EEG and ECoG data. *Clin. Neurophysiol.* 113, 43–47. doi: 10.1016/S1388-2457(01)00697-6
- Grigorev, N. A., Savosenkov, A. O., Lukoyanov, M. V., Uodoratina, A., Shusharina, N. N., Kaplan, A. Y., et al. (2021). A BCI-based Vibrotactile neurofeedback training improves motor cortical excitability during motor imagery. *IEEE Trans. Neural Syst. Rehabil. Eng.* 29, 1583–1592. doi: 10.1109/TNSRE.2021.3102304
- Guan, C. T. (2016). *Adaptation in motor imagery brain-computer interfaces and its implication in rehabilitation*. 4th international winter conference on brain-computer Interface (BCI), South Korea.
- Hashimoto, Y., and Ushiba, J. (2013). EEG-based classification of imaginary left and right foot movements using beta rebound. *Clin. Neurophysiol.* 124, 2153–2160. doi: 10.1016/j.clinph.2013.05.006
- Jurcak, V., Tsuzuki, D., and Dan, I. (2007). 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. *NeuroImage* 34, 1600–1611. doi: 10.1016/j.neuroimage.2006.09.024
- Katan, M., and Luft, A. (2018). Global burden of stroke. *Semin. Neurol.* 38, 208–211. doi: 10.1055/s-0038-1649503
- Kitahara, K., Hayashi, Y., Yano, S., and Kondo, T. (2017). Target-directed motor imagery of the lower limb enhances event-related desynchronization. *PLoS One* 12:e0184245. doi: 10.1371/journal.pone.0184245
- Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., and Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science* 297, 846–848. doi: 10.1126/science.1070311
- Kwak, N. S., Müller, K. R., and Lee, S. W. (2015). A lower limb exoskeleton control system based on steady state visual evoked potentials. *J. Neural Eng.* 12:056009. doi: 10.1088/1741-2560/12/5/056009
- Li, L., Wang, J., Xu, G., Li, M., and Xie, J. (2015). The study of object-oriented motor imagery based on EEG suppression. *PLoS One* 10:e0144256. doi: 10.1371/journal.pone.0144256
- Liu, Y.-H., Lin, L. F., Chou, C. W., Chang, Y., Hsiao, Y. T., and Hsu, W. C. (2019). Analysis of electroencephalography event-related desynchronization and synchronization induced by lower-limb stepping motor imagery. *J. Med. Biol. Eng.* 39, 54–69. doi: 10.1007/s40846-018-0379-9
- Lopez, S., Bini, F., del Percio, C., Marinuzzi, F., Celletti, C., Suppa, A., et al. (2017). Electroencephalographic sensorimotor rhythms are modulated in the acute phase following focal vibration in healthy subjects. *Neuroscience* 352, 236–248. doi: 10.1016/j.neuroscience.2017.03.015
- Machida, R., and Tanaka, H. (2018). *Visualization of ERD/ERS on leg motor imagery*. International Symposium on Affective Science and Engineering ISASE2018, pp. 1–6.
- Mane, R., Chouhan, T., and Guan, C. (2020). BCI for stroke rehabilitation: motor and beyond. *J. Neural Eng.* 17:041001. doi: 10.1088/1741-2552/aba162
- Marchesotti, S., Bassolino, M., Serino, A., Bleuler, H., and Blanke, O. (2016). Quantifying the role of motor imagery in brain-machine interfaces. *Sci. Rep.* 6:24076. doi: 10.1038/srep24076
- Meyer, S., de Bruyn, N., Lafosse, C., van Dijk, M., Michielsen, M., Thijs, L., et al. (2016). Somatosensory impairments in the upper limb Poststroke: distribution and association with motor function and visuospatial neglect. *Neurorehabil. Neural Repair* 30, 731–742. doi: 10.1177/1545968315624779
- Michael, S. A. G. (2016). Ethological action maps: a paradigm shift for the motor cortex. *Trends Cogn. Sci.* 20, 121–132. doi: 10.1016/j.tics.2015.10.008
- Naito, E. (2004). Sensing limb movements in the motor cortex: how humans sense limb movement. *Neuroscientist* 10, 73–82. doi: 10.1177/1073858403259628
- Naito, E., and Ehrsson, H. H. (2001). Kinesthetic illusion of wrist movement activates motor-related areas. *Neuroreport* 12, 3805–3809. doi: 10.1097/00001756-200112040-00041
- Naito, E., Ehrsson, H. H., Geyer, S., Zilles, K., and Roland, P. E. (1999). Illusory arm movements activate cortical motor areas: a positron emission tomography study. *J. Neurosci.* 19, 6134–6144. doi: 10.1523/JNEUROSCI.19-14-06134.1999
- Nakayashiki, K., Saeki, M., Takata, Y., Hayashi, Y., and Kondo, T. (2014). Modulation of event-related desynchronization during kinematic and kinetic hand movements. *J. Neuroeng. Rehabil.* 11:90. doi: 10.1186/1743-0003-11-90
- Neuper, C., and Pfurtscheller, G. (1996). Post-movement-synchronization of beta rhythms in the EEG over the cortical foot area in man. *Neurosci. Lett.* 216, 17–20. doi: 10.1016/0304-3940(96)12991-8
- Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., and Hallett, M. (2004). Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clin. Neurophysiol.* 115, 2292–2307. doi: 10.1016/j.clinph.2004.04.029
- Park, W., Kwon, G. H., Kim, Y. H., Lee, J. H., and Kim, L. (2016). EEG response varies with lesion location in patients with chronic stroke. *J. Neuroeng. Rehabil.* 13:21. doi: 10.1186/s12984-016-0120-2
- Pezoulas, V., Athanasiou, A., Nolte, G., Zervakis, M., Fratini, A., Fotiadis, D. I., et al. (2018). *FCLAB: An EEGLAB module for performing functional connectivity analysis on single-subject EEG data*. 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), pp. 96–99.
- Pfurtscheller, G., and da Silva, F. H. L. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* 110, 1842–1857. doi: 10.1016/S1388-2457(99)00141-8
- Pfurtscheller, G., Neuper, C., Andrew, C., and Edlinger, G. (1997). Foot and hand area mu rhythms. *Int. J. Psychophysiol.* 26, 121–135. doi: 10.1016/S0167-8760(97)00760-5
- Phon-Amnuaisuk, S. (2008). *Event-related desynchronization/synchronization of spontaneous motor actions*. 15th International Conference on Neuro-Information Processing, Auckland, New Zealand.

- Rea, M., Rana, M., Lugato, N., Terekhin, P., Gizzi, L., Brötz, D., et al. (2014). Lower limb movement preparation in chronic stroke: a pilot study toward an fNIRS-BCI for gait rehabilitation. *Neurorehabil. Neural Repair* 28, 564–575. doi: 10.1177/1545968313520410
- Ren, S., Wang, W., Hou, Z. G., Liang, X., Wang, J., and Shi, W. (2020). Enhanced motor imagery based brain-computer interface via FES and VR for lower limbs. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 1846–1855. doi: 10.1109/TNSRE.2020.3001990
- Rimbert, S., Gayraud, N., Bougrain, L., Clerc, M., and Fleck, S. (2018). Can a subjective questionnaire be used as brain-computer interface performance predictor? *Front. Hum. Neurosci.* 12:529. doi: 10.3389/fnhum.2018.00529
- Rizzolatti, G., and Sinigaglia, C. (2016). The mirror mechanism: a basic principle of brain function. *Nat. Rev. Neurosci.* 17, 757–765. doi: 10.1038/nrn.2016.135
- Romero-Laiseca, M. A., Delisle-Rodriguez, D., Cardoso, V., Gurve, D., Loterio, F., Posses Nascimento, J. H., et al. (2020). A low-cost lower-limb brain-machine interface triggered by pedaling motor imagery for post-stroke patients rehabilitation. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 988–996. doi: 10.1109/TNSRE.2020.2974056
- Rosenkranz, K., and Rothwell, J. C. (2003). Differential effect of muscle vibration on intracortical inhibitory circuits in humans. *J. Physiol.* 551, 649–660. doi: 10.1113/jphysiol.2003.043752
- Saha, S., Mamun, K. A., Ahmed, K., Mostafa, R., Naik, G. R., Darvishi, S., et al. (2021). Progress in brain computer interface: challenges and opportunities. *Front. Syst. Neurosci.* 15:578875. doi: 10.3389/fnsys.2021.578875
- Shu, X., Chen, S., Meng, J., Yao, L., Sheng, X., Jia, J., et al. (2019). Tactile stimulation improves sensorimotor rhythm-based BCI performance in stroke patients. *IEEE Trans. Biomed. Eng.* 66, 1987–1995. doi: 10.1109/TBME.2018.2882075
- Shu, X., Yao, L., Sheng, X., Zhang, D., and Zhu, X. (2017). Enhanced motor imagery-based BCI performance via tactile stimulation on unilateral hand. *Front. Hum. Neurosci.* 11:585. doi: 10.3389/fnhum.2017.00585
- Somadder, R., and Saha, D. K. (2021). *Frequency domain CSP for foot motor imagery classification using SVM for BCI application*. IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)—Leading Modern Healthcare Technology Enhancing Wellness, Electr Network.
- Song, M., and Kim, J. (2019). A paradigm to enhance motor imagery using rubber hand illusion induced by Visuo-tactile stimulus. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 477–486. doi: 10.1109/TNSRE.2019.2895029
- Storzer, L., Butz, M., Hirschmann, J., Abbasi, O., Gratkowski, M., Saupe, D., et al. (2016). Bicycling and walking are associated with different cortical oscillatory dynamics. *Front. Hum. Neurosci.* 10:61. doi: 10.3389/fnhum.2016.00061
- Takeda, K., Gomi, Y., Imai, I., Shimoda, N., Hiwatari, M., and Kato, H. (2007). Shift of motor activation areas during recovery from hemiparesis after cerebral infarction: a longitudinal study with near-infrared spectroscopy. *Neurosci. Res.* 59, 136–144. doi: 10.1016/j.neures.2007.06.1466
- Tanaka, S. (2021). Mirror neuron activity during audiovisual appreciation of opera performance. *Front. Psychol.* 11:563031. doi: 10.3389/fpsyg.2020.563031
- Tapin, A., Duclos, N. C., Jamal, K., and Duclos, C. (2021). Perception of gait motion during multiple lower-limb vibrations in young healthy individuals: a pilot study. *Exp. Brain Res.* 239, 3267–3276. doi: 10.1007/s00221-021-06199-1
- Tariq, M., Trivailo, P. M., and Simic, M. (2020). Mu-Beta event-related (de) synchronization and EEG classification of left-right foot dorsiflexion kinaesthetic motor imagery for BCI. *PLoS One* 15:e0230184. doi: 10.1371/journal.pone.0230184
- Taylor, M., Taylor, J. L., and Seizova-Cajic, T. (2017). Muscle vibration-induced illusions: review of contributing factors, taxonomy of illusions and User's guide. *Multisens. Res.* 30, 25–63. doi: 10.1163/22134808-00002544
- Triana-Guzman, N., Orjuela-Cañon, A. D., Jutinico, A. L., Mendoza-Montoya, O., and Antelis, J. M. (2022). Decoding EEG rhythms offline and online during motor imagery for standing and sitting based on a brain-computer interface. *Front. Neuroinform.* 16:961089. doi: 10.3389/fninf.2022.961089
- Vasilyev, A., Liburkina, S., Yakovlev, L., Perepelkina, O., and Kaplan, A. (2017). Assessing motor imagery in brain-computer interface training: psychological and neurophysiological correlates. *Neuropsychologia* 97, 56–65. doi: 10.1016/j.neuropsychologia.2017.02.005
- Vuckovic, A., and Osuagwu, B. A. (2013). Using a motor imagery questionnaire to estimate the performance of a brain-computer interface based on object oriented motor imagery. *Clin. Neurophysiol.* 124, 1586–1595. doi: 10.1016/j.clinph.2013.02.016
- Xu, M. P., He, F., Jung, T. P., Gu, X., and Ming, D. (2021). Current challenges for the practical application of electroencephalography-based brain-computer interfaces. *Engineering* 7, 1710–1712. doi: 10.1016/j.eng.2021.09.011
- Yang, Y. J., Jeon, E. J., Kim, J. S., and Chung, C. K. (2021). Characterization of kinesthetic motor imagery compared with visual motor imageries. *Sci. Rep.* 11:3751. doi: 10.1038/s41598-021-82241-0
- Yu, Z., Li, L., Song, J., and Lv, H. (2018). The study of visual-auditory interactions on lower limb motor imagery. *Front. Neurosci.* 12:509. doi: 10.3389/fnins.2018.00509
- Yuan, X. F., and Wang, Y. N. (2008). Parameter selection of support vector machine for function approximation based on chaos optimization. *J. Syst. Eng. Electron.* 19, 191–197. doi: 10.1016/S1004-4132(08)60066-3
- Zhang, C., Kim, Y. K., and Eskandarian, A. (2021). EEG-inception: an accurate and robust end-to-end neural network for EEG-based motor imagery classification. *J. Neural Eng.* 18:046014. doi: 10.1088/1741-2552/abed81
- Zhang, W., Song, A., Zeng, H., Xu, B., and Miao, M. (2021). Closed-loop phase-dependent vibration stimulation improves motor imagery-based brain-computer interface performance. *Front. Neurosci.* 15:638638. doi: 10.3389/fnins.2021.638638

Frontiers in Psychology

Paving the way for a greater understanding of human behavior

The most cited journal in its field, exploring psychological sciences - from clinical research to cognitive science, from imaging studies to human factors, and from animal cognition to social psychology.

Discover the latest Research Topics

[See more →](#)

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact

