Research Topics

Women in sensory neuroscience

Edited by
Monica Gori, Elena Nava and Alessia Tonelli

Coordinated by
Maria Bianca Amadeo

Published in
Frontiers in Human Neuroscience

ISSN 1664-8714
ISBN 978-2-8325-3171-6
DOI 10.3389/978-2-8325-3171-6

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

Women in sensory neuroscience

Topic editors
Monica Gori — Italian Institute of Technology (IIT), Italy
Elena Nava — University of Milano-Bicocca, Italy
Alessia Tonelli — Italian Institute of Technology (IIT), Italy

Topic coordinator
Maria Bianca Amadeo — Italian Institute of Technology (IIT), Italy

Citation
Gori, M., Nava, E., Tonelli, A., Amadeo, M. B., eds. (2023). *Women in sensory neuroscience*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3171-6

Table of contents

frontiers | Frontiers in Human Neuroscience

TYPE Editorial
PUBLISHED 24 July 2023
DOI 10.3389/fnhum.2023.1252570

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Jae Kun Shim,
University of Maryland, College Park,
United States

*CORRESPONDENCE
Alessia Tonelli
✉ tonelli.alessia@gmail.com

RECEIVED 04 July 2023
ACCEPTED 14 July 2023
PUBLISHED 24 July 2023

CITATION
Gori M, Tonelli A and Nava E (2023) Editorial: Women in sensory neuroscience.
Front. Hum. Neurosci. 17:1252570.
doi: 10.3389/fnhum.2023.1252570

Editorial: Women in sensory neuroscience

Monica Gori[1], Alessia Tonelli[1,2]* and Elena Nava[3]

[1]Unit for Visually Impaired People, Italian Institute of Technology, Genoa, Italy, [2]School of Psychology, University of Sydney, Sydney, NSW, Australia, [3]Department of Psychology, University of Milano-Bicocca, Milan, Italy

KEYWORDS

neuroscience, perception, women in science, sensory processing, sensory deprivation

Editorial on the Research Topic
Women in sensory neuroscience

Women have made remarkable contributions to this field from the early pioneers to the current generation of researchers. Their expertise and dedication have shed light on the intricate workings of the sensory systems, including vision, hearing, touch, taste, and smell. The representation of women in neuroscience has evolved, reflecting significant progress in promoting gender diversity within the field. One notable example is the groundbreaking work of Rita Levi-Montalcini, who, along with Stanley Cohen, discovered nerve growth factor (NGF), and elucidated its role in cell development and survival. Their discoveries laid the foundation for our understanding of neurotrophic factors and their implications in neural development, plasticity, and diseases. Another notable figure is May-Britt Moser, who, along with her husband Edvard Moser and their collaborator John O'Keefe, unraveled the neural mechanisms underlying spatial navigation and discovered the existence of grid cells in the brain. Their work on the brain's inner GPS earned them the Nobel Prize in Physiology or Medicine in 2014. These examples, among many others, highlight the groundbreaking contributions of women in neuroscience, pushing the boundaries of knowledge and transforming our understanding of the brain. Historically, the number of women pursuing careers in neuroscience was relatively limited, facing barriers and biases that hindered their participation. However, in recent decades, women have been actively engaged in various subfields of neuroscience, including sensory neuroscience, cognitive neuroscience, and computational neuroscience. This growth in representation has been fueled by advocacy efforts, mentorship programs, and initiatives promoting inclusivity and equal opportunities. While there is still work to be done to achieve full gender parity, the increasing number of women in neuroscience today demonstrates a positive shift toward a more diverse and inclusive scientific community supporting the importance of continuing efforts to promote gender equality and inclusivity, as it not only benefits women but the entire scientific community and society as a whole. In this Research Topic, we collect contributions from women expert on sensory neuroscience who contributed to a better understanding of important processing in the brain. We have summarized the results of these works considering three main topics that are (1) Low and high levels of sensory processing, (2) Multisensory processing, and (3) Special populations.

1. Low and high levels of sensory processing

Perceptual information can be investigated on different levels, from low-level information to more complex processes that develop across the lifespan. Regarding low-level information, Castellotti et al. investigated the role of local information in recognizing occluded images, showing that local information contributes to successfully reconstructing a visual image even when information is severely occluded. In contrast, List showed how selective attention to local and global information eliminates the level-specific priming effect in a multi-modal (audio-visual) context, while this effect has been demonstrated within both the visual and auditory modalities. However, not only low-level processes can influence perception, but also more complex processes involving body ownership and the stage at which people are in their life span. For example, La Rocca et al. found that the sense of ownership over different face identities can be influenced by congruent visuo-tactile and visuo-motor synchronous stimulation. Aloufi et al. demonstrated how sex differences can influence auditory perception. Specifically, following a literature review, the authors highlighted that women have more sensitive hearing than men. However, this higher auditory sensitivity varies concerning the menstrual cycle. Moreover, Incao et al. investigated how aging affects visual perception when faced with a spatial or temporal task, finding a general decline in perceptual acuity in both domains, but that the influence of prior knowledge determined by context does not change. However, more significant inter-individual variability is present in old age, perhaps due to different strategies that older individuals need to adopt to face higher uncertainty in the perceptual process.

2. Special population

Part of this Research Topic is dedicated to the role of the senses in promoting the typical development of different aspects of cognitive functions and how the lack of one of these can change the way blind and deaf individuals perform across tasks.

In particular, Leo et al. and Maimon et al. present evidence that congenital visual deprivation does not prevent the development of skills typically dominated by vision, such as object recognition (Leo et al.) and spatial information (Maimon et al.). Indeed, while Leo et al. show that congenitally blind individuals explore objects differently with their hands, but do not present lower accuracy with respect to late blind and sighted controls, Maimon et al. show that congenitally blind can learn to use a sensory substitution system (aimed at supporting spatial navigation) as effectively and as quickly as visually impaired individuals.

Similarly, Buyle and Crollen and Gessa et al. revealed that auditory deprivation does not impair the learning of basic and more complex skills, such as mathematics. Indeed, Buyle and Crollen found comparable subtraction and multiplication skills across deaf and hearing individuals. Gessa et al. found that sound localization in age-related hearing loss can be improved by head-movements, suggesting that self-regulation strategies and active behavior can keep spatial hearing functional.

Finally, Soker-Elimaliah et al. present a work on the relationship between pupil light reflex (PLR) and atypical neurodevelopment. PLR is associated with sensory processing and thus provides a good model to investigate the link between sensory and social functioning, especially in cases when the latter is impaired, such as autism.

3. Multimodal

We receive information about the world around us from multiple senses that interact and are combined and integrated into a multisensory framework. In the Research Topics, there is a set of important results related to multisensory processing in typical and atypical individuals considering language, reward, learning, attention, memory, and perceptual processing. In particular, Benetti et al. discussed multimodal processing in face-to-face interactions proposing a neurocognitive model of multimodal face-to-face communication psycholinguistics and sensory neuroscience. Fisher et al. showed that sensory noise might underlie attentional alterations to multisensory integration in a modality-specific manner supporting the idea that attentional paradigm might be used to study sensory processing in neurological disorders. Shvadron et al. applied multisensory knowledge to sensory substitution devices showing that by utilizing a visual-to-auditory sensory substitution device (SSD), the EyeMusic, it was possible to detect shapes by converting images to sound. Antono et al. show that visual and auditory reward cues can produce a value-driven modulation of perception. Finally, Murray and Shams review some recent findings that demonstrate a range of human learning and memory phenomena in which the interactions between visual and auditory modalities play an important role and suggest possible neural mechanisms that can underlie some recent findings.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This publication was partially supported by the MYSpace project funded by the European Research Council (ERC-grant agreement no. 948349) and the Marie Skłodowska-Curie-2021-PF-01 FLEX-U project (grant agreement no. 101064748) under the European Union's Horizon research and innovation program.

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

frontiers | Frontiers in Human Neuroscience

TYPE Original Research
PUBLISHED 08 September 2022
DOI 10.3389/fnhum.2022.941593

OPEN ACCESS

EDITED BY
Sieu Kieu Khuu,
University of New South Wales,
Australia

REVIEWED BY
Lore Thaler,
Durham University, United Kingdom
Magdalena Szubielska,
The John Paul II Catholic University
of Lublin, Poland

*CORRESPONDENCE
Fabrizio Leo
fabrizio.leo@iit.it

SPECIALTY SECTION
This article was submitted to
Sensory Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 11 May 2022
ACCEPTED 16 August 2022
PUBLISHED 08 September 2022

CITATION
Leo F, Gori M and Sciutti A (2022) Early blindness modulates haptic object recognition.
Front. Hum. Neurosci. 16:941593.
doi: 10.3389/fnhum.2022.941593

Early blindness modulates haptic object recognition

Fabrizio Leo[1]*, Monica Gori[2] and Alessandra Sciutti[1]

[1]Cognitive Architecture for Collaborative Technologies Unit, Istituto Italiano di Tecnologia, Genova, Italy, [2]Unit for Visually Impaired People, Istituto Italiano di Tecnologia, Genova, Italy

Haptic object recognition is usually an efficient process although slower and less accurate than its visual counterpart. The early loss of vision imposes a greater reliance on haptic perception for recognition compared to the sighted. Therefore, we may expect that congenitally blind persons could recognize objects through touch more quickly and accurately than late blind or sighted people. However, the literature provided mixed results. Furthermore, most of the studies on haptic object recognition focused on performance, devoting little attention to the exploration procedures that conducted to that performance. In this study, we used iCube, an instrumented cube recording its orientation in space as well as the location of the points of contact on its faces. Three groups of congenitally blind, late blind and age and gender-matched blindfolded sighted participants were asked to explore the cube faces where little pins were positioned in varying number. Participants were required to explore the cube twice, reporting whether the cube was the same or it differed in pins disposition. Results showed that recognition accuracy was not modulated by the level of visual ability. However, congenitally blind touched more cells simultaneously while exploring the faces and changed more the pattern of touched cells from one recording sample to the next than late blind and sighted. Furthermore, the number of simultaneously touched cells negatively correlated with exploration duration. These findings indicate that early blindness shapes haptic exploration of objects that can be held in hands.

KEYWORDS

haptics, object recognition, blindness, exploration strategies, perception and action

Introduction

Humans can visually recognize objects in complex scenes in about one-tenth of a second (Potter, 1976; Thorpe et al., 1996). However, objects recognition is not a prerogative of vision. For instance, we can accurately identify real objects using only touch, although with a slower recognition time, in the order of seconds (Klatzky et al., 1985). The difference in recognition time between vision and touch is also due to intrinsic differences between the two sensory systems. Vision is usually characterized by holistic acquisition of information, whereas, touch often encodes information in a more sequential, and slower, fashion (Cattaneo and Vecchi, 2008). For instance, vision can decode simultaneously attributes of objects such as color and shape whereas touch may need different exploratory procedures, applied in sequence, to detect object properties such as texture and shape. We use indeed lateral motion to assess texture and contour-following to identify the shape (Lederman and Klatzky, 1987; Klatzky and Lederman, 1992). Visual and haptic object perception also differs for the weight they assign to different object properties (Lacey and Sathian, 2014). For instance, shape is more important than texture when visually categorizing, whereas shape and texture are approximately equally weighted in haptic categorization (Cooke et al., 2007).

However, visual and haptic object perception also shares some properties. For example, when considering object categorization, both vision and haptics show categorical perception, i.e., discriminability increases markedly when objects belong to different categories and decrease when they belong to the same category (Gaißert et al., 2012). In addition, both sensory modalities seem to be viewpoint-specific, i.e., they best recognize an object when it is oriented in a specific way although vision prefers "front-view" and haptics prefer "back-view" orientation (Newell et al., 2001).

Scientific works support the idea that these similarities may also have a neurophysiological foundation. Indeed, the visual and tactile sensory systems share some analogies also at the neural level (Amedi et al., 2005). They are both characterized by a hierarchical organization of increasing complexity. For instance, the unspecific tactile input is firstly processed in areas 3b and 1 of the primary somatosensory cortex, then by area 2 which shows selectivity to attributes of objects such as curvature and, finally, by the anterior intraparietal sulcus (IPS), which shows preference to overall shape rather than primitive attributes such as curvature (Bodegård et al., 2001). Both visual and tactile sensory systems show a topographical organization, i.e., adjacent parts of the space are mapped in adjacent parts in retinotopic and somatotopic cortical maps. More importantly, vision and touch may activate similar brain areas when exploring objects, for example, the visual ventral and dorsal pathways are also involved during similar haptic tasks (Amedi et al., 2005; Lacey and Sathian, 2011). For instance, James et al. (2002) found that haptic object exploration activated the middle and lateral occipital areas active in the corresponding visual exploration task. These cortical areas may be part of a network of neural substrates responsible for a supramodal representation of spatial information (Cattaneo and Vecchi, 2008; Loomis et al., 2013; Ottink et al., 2021). The existence of such supramodal representation is also suggested by other findings. For instance, Giudice et al. (2011) showed similar biases and updating performance when learning visual or tactile maps.

One might wonder what happens when the visual cortex does not receive visual input, as in blindness. It has been shown how the visual cortex can be functionally reprogramed in the blind to process tactile [see Sathian and Stilla (2010) for a review] or auditory stimuli (Kujala et al., 1995; Burton, 2003; Campus et al., 2019). As a consequence, the overall cortical representation of the tactile sense may be larger in the blind relative to sighted persons which may help explaining some superior tactile abilities, such as the higher tactile acuity, in the former population (Penfield and Boldrey, 1937; Goldreich and Kanics, 2003; Bliss et al., 2004; Wan et al., 2010; Norman and Bartholomew, 2011; Wong et al., 2011). However, haptic object recognition is a complex skill involving not only low-level tactile processing but also motor, memory, and spatial components. In particular, it has been suggested that visual mediation, that is, the translation of the tactile input into a visual image, may enhance haptic object recognition (Lederman et al., 1990). Therefore, according to the visual mediation hypothesis, we may hypothesize that object recognition based only on haptics may be superior in the late blind relative to congenitally blind or blindfolded sighted controls. Late blind individuals may indeed benefit of both extended haptic practice and the ability to translate the haptic information into a visual representation since they had seen earlier in life. Other researchers suggested that visual mediation may conduct to another advantage, that is the ability to represent spatial information in allocentric perspective. With allocentric representation we mean the ability to code spatial information based on an external perspective, independent from the observer, whereas, a representation is egocentric when it is based on the perspective of the observer (Taylor and Tversky, 1992). Allocentric representations are usually associated with higher spatial performance (Lawton, 1994; Meneghetti et al., 2011). It has been shown how blind individuals might prefer egocentric representations of spatial information while sighted persons tend to code the same information as allocentric, at least in the context of learning maps of environments (e.g., Noordzij et al., 2006). Toroj and Szubielska (2011) applied this framework to explain why their late blind participants, using an allocentric strategy when visualizing object shapes in their imagery, better identified such shapes than congenitally blind. The differentiation between egocentric and allocentric leads to the hypothesis that object recognition may depend also on the orientation of the objects relative to the participant. For instance, it has been shown how object recognition is impaired when the object is rotated with respect to the orientation of the learning phase which may be interpreted with the difficulty of moving from an egocentric to an allocentric perspective. This performance degradation is visible in the sighted regardless of the sense involved in recognition, that is vision or touch (Lacey et al., 2007). On the contrary, Occelli et al. (2016) showed that in the congenitally blind object recognition is view-independent, that is accuracy is not affected by the rotation of the learned object. Another result of this study is that overall no difference in performance between blind and sighted was observed. Szubielska and Zabielska-Mendyk (2018) also found similar ability in mentally rotating tactile figures in congenitally blind and sighted individuals.

Another line of research used two dimensional depictions of 3D shapes presented on raised line drawings. Using this kind of material, Heller (1989) found better recognition performance in late blind compared to sighted or congenitally blind persons. These latter two groups showed similar performance. On the contrary, Lederman et al. (1990) found that congenitally blind did worse than sighted in haptic recognition of 3D shapes and Gori et al. (2010) showed that congenitally blind children had higher orientation discrimination threshold compared to age matched controls. Collectively, these findings have been interpreted in terms of the necessity to visually translate the haptic information. In this perspective, the better performance in late blind may be the result of two factors: (1) their well-trained tactile skills; (2) their possibility to visually translate haptic information thanks to the fact they had seen earlier in life. This latter hypothesis is also well in line with a previous finding showing how the lack of visual experience in the early years of life can disrupt spatial processing in other sensory modalities (i.e., audition) suggesting the idea the visual system calibrates auditory spatial maps (Gori et al., 2014). However, the limited performance in early blind may not be present when manipulating real tridimensional objects. An early attempt to investigate this behavior in sighted and congenitally blind children has been performed by Morrongiello et al. (1994). The authors failed to find any difference in performance between the two populations. However, more recently, Norman and Bartholomew (2011) found even superior recognition accuracy of 3D shapes, not resembling daily-life objects, in early and late blind, but not in congenitally blind compared to sighted. Certainly, the contradiction between the studies may be due to the different tasks used and to possible differences in the tested populations.

In addition, to the best of our knowledge, most studies on this topic devoted little attention to the haptic patterns of exploration. For instance, studies using raised-lines drawings or textured pictures mainly focused on the final outcome in performance, that is recognition accuracy and time without investigating the haptic behavior conducting to that performance (e.g., Heller, 2002; Picard and Lebaz, 2012; Vinter et al., 2020). In Morrongiello et al. (1994), the authors also analyzed some basic haptic strategies of children exploring 3D objects. For instance, they measured the number of unique parts composing the object that was touched in a trial or the number of repetitions of exploration of those unique parts by examining video recordings. However, using this method, finer exploration features such as the number of touches of unique parts, their temporal frequency or the way subjects manipulated and rotated the objects could not be examined. Such haptic patterns may provide interesting complementary information as Leo et al. (2022) showed that different outcomes in performance in a haptic task may be associated with different haptic exploration strategies. Similarly, accuracy in haptic spatial tasks has been shown to depend on the level of development: children under 9 years of age showed indeed less effective haptic exploration than adults (Sciutti and Sandini, 2020). Furthermore, investigating such more detailed haptic exploration strategies may be necessary for identifying differences between groups of persons differing in spatial and visual ability. Therefore, in our study, we aimed at investigating: (1) how the performance in a haptic object recognition task is influenced by the level of visual ability; (2) how the level of visual ability shapes haptic exploration patterns. To do so, early blind, late blind, and sighted participants performed a haptic recognition task using an instrumented cube that measures the touches on its faces as well as its rotation, that is, the iCube (Sciutti and Sandini, 2019; Sciutti et al., 2019). As in Sciutti et al. (2019), we attached small pins on cube faces in varying number and asked participants to explore the cube twice, with the task of understanding whether any change occurred in the pins distribution between the first and the second presentation. This design is similar to a "study-test" paradigm to assess memory and recall (Pensky et al., 2008). Our study has a data-driven exploratory nature and several dependent variables recorded by iCube have never been collected in visually impaired subjects. However, we could at least expect that: (1) recognition accuracy may be similar across groups since the simple cube-like shape should not favor participants able to take advantage of a visual-mediation strategy; (2) both congenitally and late blind participants might be faster in doing the haptic task since they have larger haptic experience; (3) if it is true that blind persons and, particularly, congenitally blind prefer an egocentric representation of spatial information they might tend to rotate less the cube while exploring to facilitate the association of each cube face to its relative orientation.

Materials and methods

Participants

A group of congenitally blind (CB, $n = 7$, four females), a group of late blind (LB, $n = 10$, five females) and a sighted control group, age and gender matched with the visually impaired groups (SI, $n = 16$, nine females), took part in the study (see **Table 1**). One congenitally blind was excluded due to a technical issue with data collection. Following the World Health Organization (WHO) guidelines, we defined blindness as vision in a person's best eye with correction of less than 20/500 or a visual field of less than 10°. All LB lose sight after 6 years of age. CB age ranged from 23 to 49 years (mean age = 35; SD = 9.5). LB age ranged from 30 to 61 years (mean age = 43.9; SD = 12). SI age ranged from 22 to 64 years (mean age = 40.7; SD = 12.1).

TABLE 1 Characteristics of the blind participants.

Participant	Gender	Age (years)	Etiology of visual impairment	Age at onset of complete blindness	Residual vision
Congenitally blind					
cb01	F	34	Retinopathy of prematurity	Birth	None
cb02	F	23	Retinopathy of prematurity	Birth	Light and shadow
cb03	M	32	Retinitis pigmentosa	Birth	None
cb04	M	29	Leber amaurosis	Birth	None
cb05	F	49	Retinopathy and glaucoma	Birth	Light and shadow
cb06	F	43	Atrophy optic nerve	Birth	None
Late blind					
lb01	M	34	Macular degeneration	20	Light and shadow, 1% visual field
lb02	F	56	Retinitis pigmentosa	35	Light and shadow
lb03	M	34	Corneal opacity	17	None
lb04	F	44	Accident, loss of retina	18	Light and shadow
lb05	F	61	Retinitis pigmentosa	40	Light and shadow
lb06	M	30	Leber amaurosis	19	Light and shadow
lb07	F	31	Optic nerve tumor	6	Light and shadow
lb08	F	61	Uveitis	11	None
lb09	M	45	Retinitis pigmentosa	34	Light and shadow
lb10	M	43	Retinitis pigmentosa	26	Light and shadow

Participants reported no conditions affecting tactile perception, or cognitive impairment. Blind participants were selected by the Istituto David Chiossone in Genoa and by the UVIP Unit of the Istituto Italiano di Tecnologia and agreed to participate on a voluntary basis. The experimental protocol was approved by the ethics committee of the local health service (Comitato Etico Regione Liguria, Genoa, Italy; Prot. IIT_UVIP_COMP_2019 N. 02/2020, 4 July 2020). All participants provided their written informed consent.

The iCube

The iCube (v3) is an instrumented cube designed at IIT which measures its orientation in space as well as the location of contacts on its faces. This information is conveyed wirelessly to a laptop. iCube is of about 5 cm side, it has 16 cells per face and a weight of about 150 g (see **Figure 1**). Touch sensing is based on a 4 × 4 array of Capacitive Button Controllers (CY8CMBR2016) developed by Cypress Semiconductor Corporation. These are based on Multi Touch technology, allowing detection of simultaneous touches and support up to 16 capacitive cells (6 mm × 6 mm × 0.6 mm), which could be organized in any geometrical format, e.g., in matrix form. Each face of iCube is made with one of these boards. Their sensitivity, i.e., the smallest increase in capacitance that could be detected clearly as a signal, is set to 0.3 pF to allow the device to sense contacts without the need to apply pressure. Spatial orientation of the cube is estimated by a Motion Processing Unit^{TM} (MPU), a nine axes integrated device, combining a three axes MEMS gyro, a three axes MEMS accelerometer, a three axes MEMS magnetometer and Digital Motion Processor^{TM} (DMP). The MPU combines information about acceleration, rotation and gravitational field in a single flow of data. Data from iCube are sent to a laptop through a serial protocol. The transmission is performed through a radio module NRF24L01 (Nordic Semiconductor, Trondheim, Norway). The firmware of the device is designed to maximize the speed of capture of information from the boards measuring touches. The acquisition is always as fast as possible: faster when least faces are touched simultaneously and slower when it needs to encode information from multiple faces. As a result, the average sampling rate of the device was about 5 Hz (i.e., one sample every 203 ± 113 ms, SD). As in Leo et al. (2022), data were subsequently interpolated to analyze the temporal evolution of exploration at a constant temporal rate. Data generated in this study was further analyzed in Python (Python Software Foundation) to extract the pattern of touches, the amount of iCube rotation and the speed of rotation (see Section "Data Analysis").

Procedure

The experimenter positioned on iCube faces a set of raised plastic pins (diameter: 0.3 cm, height: 0.2 cm). Each face contained from 0 to 5 pins with no limitation of the presence of two or more equal faces. The participant was seated in front of a table, where the iCube was positioned on a support.

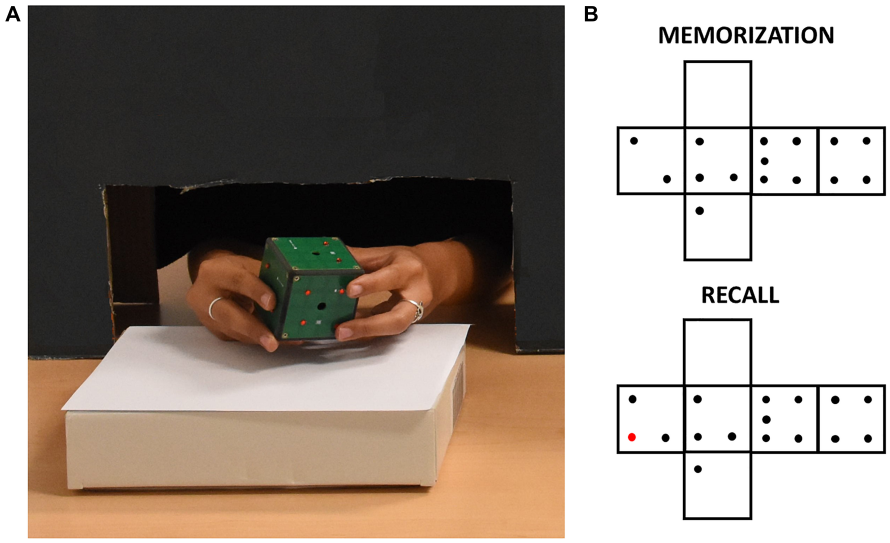


FIGURE 1
(A) A participant exploring the iCube with raised red pins positioned on its faces. The black cardboard panel avoided visual inspection of the device while allowing unconstrained haptic exploration. **(B)** Example of pins configurations of one trial. The red pin in the recall configuration indicates the difference with respect to the memorization configuration.

Whenever a sighted participant was tested, a cardboard panel was placed on the table between him/her and the cube to avoid any visual inspection of the device. To do so, a black curtain was also fixed to the lower part of the panel on the side of the participant. This panel allowed anyway comfortable movements of participants' upper limbs (see **Figure 1**). Before the experiment, participants performed a familiarization phase. In this phase, they first explored the cube without pins for a few seconds to get acquainted with it. After that, they did two practice trials in which they familiarized themselves with the experimental task, i.e., they were asked to explore the cube twice trying to understand whether any change occurred in the pins allocation between the first (memorization) and the second exploration (recall). Particularly, they were asked to report whether the cube in the second exploration was the "same" or "different" compared to the cube in the first exploration. When participants had proven to understand the task, the real experiment began. They did three trials in sequence for a total of six cube explorations for each participant. Between the memorization and recall phases, the cube could remain the same, but rotated on the support, or could be changed (e.g., by removing or adding one pin to one of the faces, see **Figure 1** for an example). The experimenter rapidly operated these changes, with an interval between explorations lasting on average less than a minute. We opted for two "different" and one "same" trial to minimize participants' fatigue as the latter trial-type has been shown as more difficult in previous studies (Norman et al., 2004; Sciutti et al., 2019). The experiment lasted about 30 min on average, including explanations and cube preparation.

Data analysis

Data about touches and rotations recorded by iCube were processed in Python following the methods used in Leo et al. (2022) and briefly described below.

Touches

The cube reported for each timestamp a tactile map, i.e., a list of 16 elements of zeros and ones, where one represents a touched cell. These tactile maps were independently interpolated at a constant rate of 0.2 s, i.e., a value close to the average sample rate of the device. We then spatiotemporally filtered the tactile maps to select the explorative touches, i.e., touches directly related to the exploration of a face to detect and count its pins, from the holding touches, i.e., touches that only reflect the holding or support of the device. This filter was based on simple matching coefficient (SMC: $\frac{number\ of\ matching\ attributes}{number\ of\ attributes} = \frac{M00+M11}{M00+M01+M10+M11}$) which is a measure of similarity of samples sets with scores between 0 and 1, where 1 indicates perfect similarity and 0 indicates perfect diversity. M_{11} is the total number of cells where sample 1 and sample 2 both have a value of 1 (active); M_{01} is the total number of cells where the status of sample 1 is 0 (inactive) and the status of sample 2 is 1 (active); M_{10} is the total number of cells where the status of sample 1 is 1 (active) and the status of sample 2 is 0 (inactive); M_{00} is the total number of cells where sample 1 and sample 2 both have a value of 0 (inactive). Then, as in Leo et al. (2022) we assumed that explorative touches were characterized by higher variability in space and time than holding touches. Holding touches, by definition, are indeed stable in time to allow a secure grasping and movement of objects. For instance, the lateral motion exploratory procedure often associated with active exploration of a surface's tactile features such as texture is characterized by highly dynamic movement of the hand in contact with the object. This kind of movement would translate for our sensors in a rapid change of status of cells activation in a face, resulting in lower SMC for consecutive temporal samples. Therefore, at each time interval we only considered explorative touches those measured on the face with the lowest SMC computed concerning the previous sample. If more than one face shared the lowest SMC, we considered the touches of all those faces, unless the SMC was 1 for all faces which would likely indicate the cube lying untouched on the table. We then computed the mean SMC of the explored faces for each trial. We used this variable as an indirect measure of velocity in exploring a face since, for instance, a very low SMC between two consecutive samples (0.2 s duration each) means that the participant touched very different cells between the two samples. We also computed: (1) the exploration duration of each trial as the time between the first and last touch of the participant (via manual cutting for each file the initial and final phases of recording, when less than two cells were active); (2) the mean exploration duration for each face; (3) the variability (i.e., standard deviation) of the mean exploration duration for each face; (4) the touch frequency, i.e., the number of touches per time unit (s); (5) the mean number of active cells per sample in the explored faces (after removing samples with no active cells).

Rotations

The information about the orientation of iCube in time was provided in the form of quaternions. Quaternions were interpolated at a constant sample rate of 0.2 s *via* spherical linear interpolation (SLERP). Then, we computed the instantaneous angular variation by measuring the angle traversed over time by each of the three unitary axes orthogonal to the faces of iCube. In particular, given one axis:

$$\Delta \text{angle}_{\text{axis}}(\text{t}) = \arctan\left(\left|\frac{axis\,(t)\,xaxis\,(t-1)}{axis\,(t)\cdot axis\,(t-1)}\right|\right) * 180^\circ/\pi \quad (1)$$

We integrated over time the rotations performed by the three axes to estimate the rotation impressed to iCube in all the possible directions. To quantify the amount of rotation, we considered the maximum value among cumulative sums of the rotations executed by the three axes. The instantaneous rotation speed was instead computed by dividing $\Delta\text{angle}_{\text{axis}}(\text{t})$ for its time interval (i.e., 0.2 s) and averaging the results across the three axes and all the instants in a trial in which iCube was in motion (i.e., angular velocity $> 1^\circ/\text{s}$). As in Sciutti et al. (2019), this selection was made to assess the actual velocity of rotation when the rotations were executed, without spuriously reducing the estimate with the analysis of the static phases. In addition, we determined for each timepoint the absolute and relative orientation of each face of iCube. With absolute orientation we mean the cardinal direction of the normal of a face (with labels such as "North," "East," etc.). With relative orientation of a face we mean its orientation in the participant's perspective (with labels such as "up," "rear," etc.). See Leo et al. (2022) for more details about these estimations.

Transition matrices

We computed the transition matrices for all the trials of the experiment, i.e., six by six matrices in which each cell corresponds to the percentage of cases in which the transition has occurred between the face individuated by the row number and the face corresponding to the column number (for instance, from "front" to "left"). Each trial is indeed characterized by a temporal sequence of explored faces (e.g., left, up, front, left, etc.). The transition matrix is computed by counting and summing the number of transitions (e.g., from "left" to "up") and converting these numbers into percentage of occurrences. In particular, we computed a transition matrix for each trial in each participant (i.e., three matrices for the "memorization" trial type and three matrices for the "recall" trial type). Then, for each transition matrix we computed two different scores (Leo et al., 2022): (1) the maximum diagonal score; (2) the mean number of different transitions. The maximum diagonal score is the highest value in the diagonal cells. These cells reflect the tendency to select specific relative orientations as objects of spatial attention (e.g., a high proportion in the "from right to right" cell indicates that participant preferentially explored the rightward face and rotated the cube to position the face they wanted to explore toward their right). The number of different transitions is a measure of exploration variability (e.g., low numbers indicate participants selected less orientations to explore, i.e., less variability). For instance, a participant with a high maximum diagonal score and a low number of different transitions would be characterized by a very focused and systematic exploration reflecting high spatial ability (Leo et al., 2022). Finally, we measured the number of returns to already explored faces. For this measure, we did not consider the sequence of explored orientations but the sequence of explored faces in terms of their label (from 1 to 6). This measure may be relevant because a previous study showed that participants with lower spatial skill showed also an higher number of returns (Leo et al., 2022).

Statistical analyses

Statistical analyses were performed using R. To sum up, we analyzed the following dependent variables: (1) recognition accuracy; (2) exploration duration (in s); (3) number of touches; (4) touch frequency (touches/s); (5) amount of rotation (°); (6) rotation velocity (°/s); (7) maximum diagonal score; (8) number of different transitions; (9) exploration duration per face; (10) variability of exploration duration per face; (11) number of returns; (12) mean number of active cells per sample; (13) mean SMC. The independent variables were the Group (early-blind, late-blind, sighted) and Trial Type (memorization vs. recall). Since we did not have specific hypotheses regarding the interaction between Group and Trial Type and since the comparison between memorization and recall in the same task has been already investigated in Sciutti et al. (2019) we only focused on group differences. Given the high number of dependent variables we ran an explorative MANOVA including all the normally distributed dependent variables (all but recognition accuracy) with Group as between factor. For recognition accuracy, after a Box-Cox transformation using the MASS R package (Venables and Ripley, 2002), we estimated a Bayes factor to compare the fit of the data under the null hypothesis and the alternative hypothesis using BayesFactor R package (Morey and Rouder, 2011). Data normality was assessed with Shapiro-Wilk tests. After the MANOVA we also performed a Linear Discriminant Analysis (LDA) as follow-up with the goal of defining which linear combination of dependent variables led to maximal group separability. We then conducted univariate ANOVA on the dependent variables that showed higher coefficients in the LDA followed by *t*-tests as *post hoc*. We corrected for multiple comparisons using Benjamini/Hochberg FDR correction (Benjamini and Hochberg, 1995a,b). We set statistical significance at $p < 0.05$.

Results

As for the iCube recognition, the mean accuracy was 72% for the CB, 77% for the LB and 69% for the SI. The estimated Bayes factor suggested that the data were 3.7 times more likely to occur under a model without including an effect of group, rather than a model with it.

The MANOVA revealed a significant difference between groups in the haptic exploration variables [Roy (2,12) = 1.84, $p = 0.018$].

The follow-up LDA identified two linear discriminants which accounted for a percentage of separation between groups of 74.8 and 25.2%, respectively. The haptic variables which were able to discriminate more strongly the groups were the mean SMC, the mean active cells per sample and the maximum diagonal score. **Table 2** shows the normalized coefficients of linear discriminants. **Figure 2** shows participants distribution along the two discriminants. It is evident how the three groups concentrate in different areas defined by the two discriminants. Both CB and LB participants tend to have higher scores than SI in LD1. As for the LD2, while SI showed intermediate levels, LB and CB showed higher and lower scores, respectively. Finally, CB tend to form a quite separate cluster whereas LB and SI clusters show higher superposition.

TABLE 2 Coefficients of linear discriminants (LDA).

Haptic variable	LD1	LD2
Exploration duration	0.09	0.06
Number of touches	−0.02	−0.01
Touch frequency	−0.42	0.59
Amount of rotation	0.00	0.00
Rotation velocity	−0.07	0.00
Maximum diagonal score	**6.56**	**−1.20**
Number of different transitions	0.48	0.12
Exploration duration per face	−0.45	−0.69
Variability in exploration duration per face	0.49	0.87
Number of returns	−0.18	0.34
Mean active cells per sample	**3.17**	**−1.44**
Mean SMC	**−9.36**	**15.51**

Coefficients for each linear discriminant. Bold indicates haptic variables whose linear combination discriminated more strongly between groups (absolute value > 1).

In order to statistically substantiate these differences, we ran a one-way ANOVA for each of the three haptic variables that contributed more in discriminating the groups, i.e., max diagonal score, mean SMC and mean active cells per sample. As for the maximum diagonal score, the groups did not differ [CB = 3.41, LB = 3.27, SI = 2.24; $F_{(2,29)} = 0.87$, $p = 0.43$]. As for the mean active cells per sample in the explored face, the groups tend to differ [CB = 5.24, LB = 4.32, SI = 4.14; $F_{(2,29)} = 3.75$, $p_{unc} = 0.035$, $p_{fdr} = 0.07$]. *Post hoc* tests showed that the number of active sensors was higher in the CB than in the SI [$t_{(44.8)} = -4.96$, $p_{fdr} < 0.001$; see **Figure 3A**] and in the LB [$t_{(55.5)} = 3.91$, $p_{fdr} = 0.00038$; see **Figure 3A**]. The comparison between SI and LB was not significant ($p = 0.22$). As for the mean SMC, this score tend to differ in the three groups [CB = 0.77, LB = 0.81, SI = 0.80; $F_{(2,29)} = 3.38$, $p_{unc} = 0.047$, $p_{fdr} = 0.07$; see **Figure 3B**] since it was lower in the CB than in the SI [$t_{(59.6)} = 4.14$, $p_{fdr} = 0.00017$] and in LB [$t_{(68.6)} = -4.31$, $p_{fdr} = 0.00016$]. No difference was observed between LB and SI ($p = 0.58$).

A lower SMC and higher mean number of active cells per sample in the explored faces are potentially indexes of faster exploration because the former indicates the participant considerably changed the touched cells from one sample to the next and the latter shows that more cells were simultaneously considered. Therefore, we further hypothesized that SMC score and number of active cells per sample would correlate positively and negatively, respectively, with exploration duration. To verify these hypotheses, we computed Pearson's correlation coefficients (r). Results showed that the SMC did not correlate with exploration duration ($r = 0.21$, $p = 0.127$, one-tailed), whereas, the number of active cells per sample did ($r = -0.38$, $p_{fdr} = 0.03$, one-tailed; see **Figure 4**).

Discussion

Our study had two different aims: first, investigating whether the level of visual ability modulates haptic object

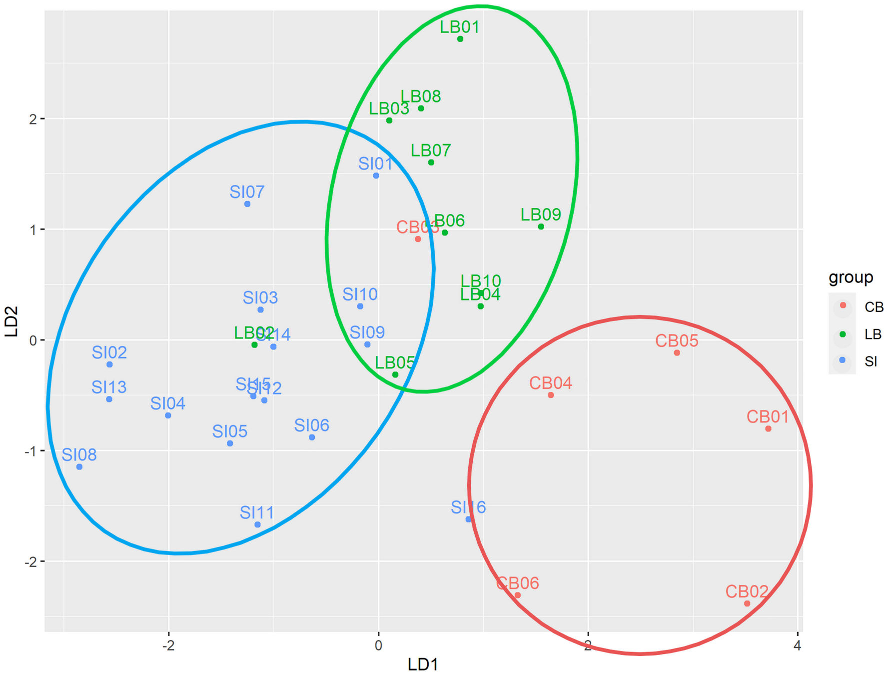


FIGURE 2
Scatterplot of participants distribution in the two LDA dimensions. The diagram depicts congenitally blind (CB) as red circles, late blind (LB) as green circles and sighted controls (SI) as blue circles. The labels above each circle specify participants' code. Ellipses indicate the three identified clusters. Note as the three groups tend to concentrate in different areas of the 2D space as defined by the two discriminants.

recognition; second, highlighting possible differences in the exploration strategies in congenitally blind, late blind, and sighted individuals using a sensorized cube. To do so, we asked a group of congenitally blind, a group of late blind and a group of sighted persons (who could not see the device) to explore twice an iCube with pins attached to its faces. In the second exploration, the iCube could have the same pins disposition, although the cube would be presented in a different orientation, or a small change in pins disposition, e.g., one pin less or more in one of the faces. Participants had to report whether the two presented cubes had the same pin disposition, or they differed. The main advantage of using the iCube compared to common daily-life objects lies in that it allows a free and unconstrained manipulation while keeping the possibility of accurately measuring how it is touched and its orientation in space without the need to use video recordings.

Our results showed that the level of visual ability does not influence the accuracy in recognizing the cube. This finding is in line with Morrongiello et al. (1994), who, in addition, also failed to observe differences between blind and sighted children in terms of exploration behavior. However, in our case, we showed evidence of different haptic strategies between congenitally blind and the other groups. Indeed, congenitally blind tend to touch simultaneously more cells in each recording sample when exploring a face than late blind and sighted persons, suggesting that they learnt to consider a larger tactile space with a single touch. They also tend to change touched cells more quickly than the other groups. This is an important result because it suggests that congenitally blind persons may have a peculiar way to explore the environment through touch, which differentiates them even from late blind persons characterized by many years of complete blindness, as in our sample of participants. Furthermore, we observed that the number of simultaneously touched cells negatively correlated with exploration duration. If we can cover a larger tactile space with a single touch, then the time needed to fully explore an object decreases. It should be noted that a previous study showed evidence of an impairment in haptic recognition of faces in the congenitally blind and not in late blind suggesting that early visual experience is necessary to process face features (Wallraven and Dopjans, 2013). However, there is also evidence that faces may be special kind of "objects" processed by dedicated brain areas in the human visual system, such as the fusiform gyrus (Puce et al., 1995; Yue et al., 2006). Therefore, findings on faces recognition in the blind may not be easily translated to different types of objects.

Our third hypothesis, i.e., blind participants would rotate less the cube was not supported by results. However, this may simply be due to the reduced power of our analysis since congenitally blind and late blind tended to rotate less the device (560° and 517°, respectively) than sighted (710°).

Importantly, our findings do not seem to be due to differences in spatial memory in the groups of participants.

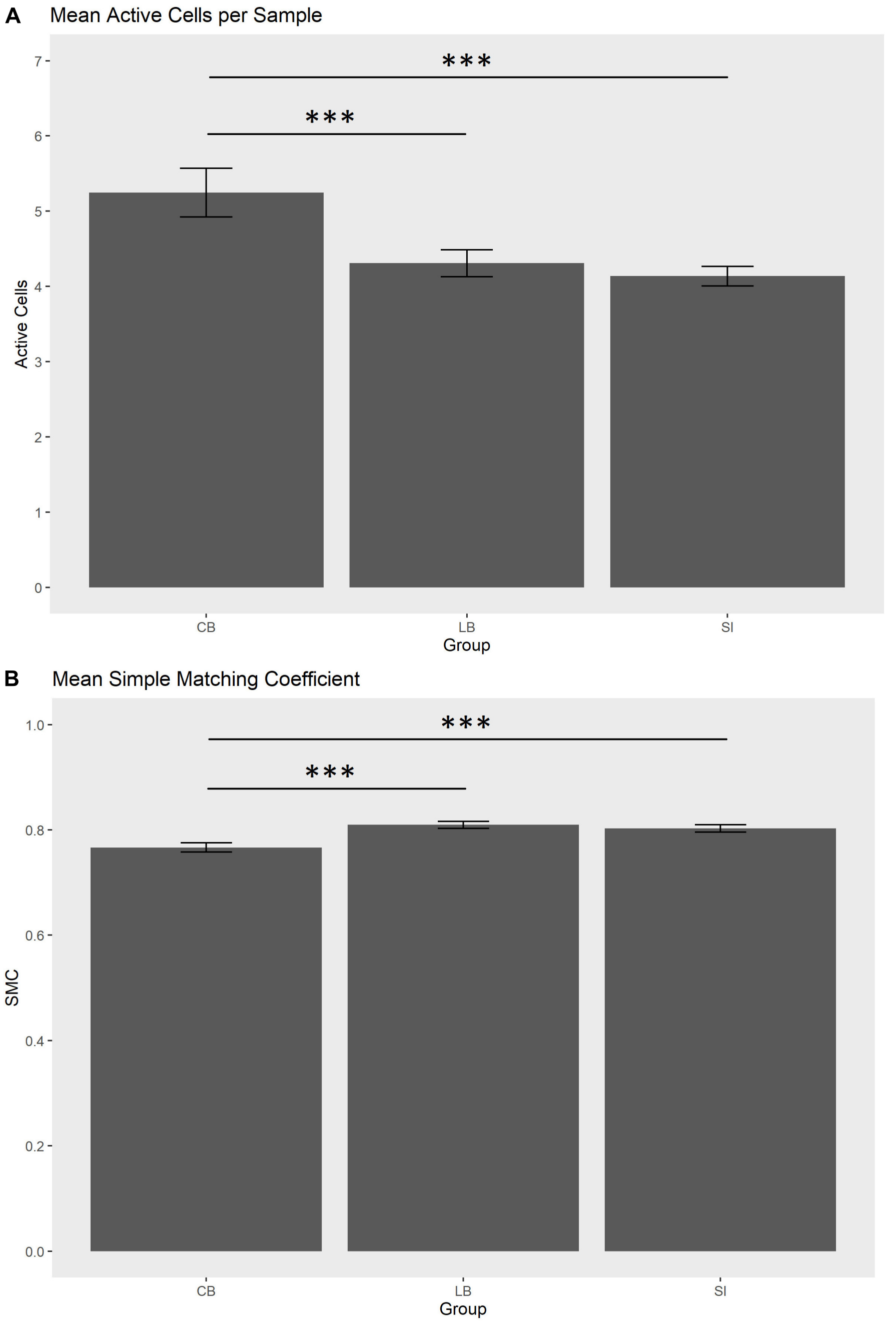


FIGURE 3
(A) Mean active cells per sample (0.2 s) in the explored face. **(B)** Mean simple matching coefficient (SMC) in the explored face. Whiskers indicate SEM. ***$p_{fdr} < 0.001$.

There is evidence that congenitally blind subjects may have difficulties in specific spatial memory tasks, particularly when they have to memorize and recall two separate haptic spatial configurations (Vecchi et al., 2004; Leo et al., 2018, 2020) or sequences of semantic sounds. However, in our study the congenitally blind showed a similar recalling accuracy than the other groups. Our task did not impose indeed a heavy burden on spatial memory since participants were required to keep in memory only five items (the number of pins in five faces) and their relative location. On the contrary, in Leo et al. (2018) participants had to memorize an average of 2.5 targets randomly located in a 3 × 3 grid and they had to do so for two different grids presented in sequence. This task is much more complex because there are many ways to place 2.5 targets in

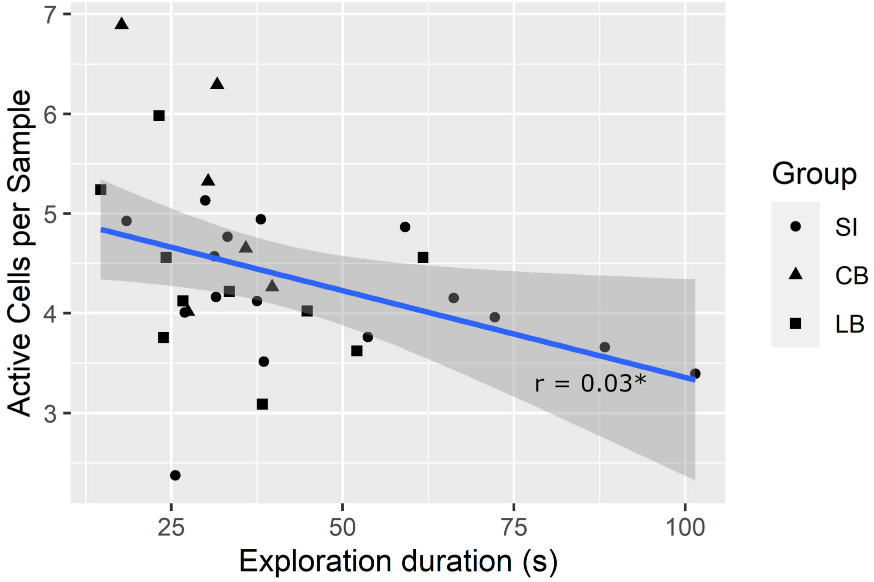


FIGURE 4
Correlation between exploration duration and mean number of active cells per sample. $*p_{fdr} < 0.05$.

a nine-elements grid and participants had to keep in memory two of these grids.

The conflict between our and Morrongiello et al.'s (1994) findings who did not observe haptic differences in object recognition between blind and sighted participants may be due to several reasons: (a) Morrongiello and coauthors tested only children. It is possible that the differences we found in haptic patterns would emerge only later in life, as a consequence of the more extended haptic training [but see Withagen et al. (2012) for a similar result with adults]; (b) they used common daily-life objects, whereas we used two cubes eventually differing between each other only for relative pin disposition on the surface of their faces; (c) they studied haptic behavior through evaluation of video recordings, that is with a methodology and a selection of dependent variables which may be not sensitive enough to detect subtle differences in exploration procedures.

On the other hand, there is also evidence in the literature regarding differences in exploratory procedures between blind and sighted children, although in studies using different materials and methods. For instance, Vinter et al. (2012) asked blind, low vision, and blindfolded sighted children to haptically explore raised-line drawings whose comprehension was subsequently evaluated through drawings of the remembered shapes. Briefly here, results showed how blind children used more types of exploratory procedures, as defined in Davidson (1972), Lederman and Klatzky (1987, 1993), and Wijntjes et al. (2008), than their sighted peers. The use of certain kinds of procedures (e.g., contour following) also correlated with drawing performance. However, this study referred to the classical exploratory procedures originated by the seminal work of Lederman and Klatzky (1987) which cannot easily be translated to the case of solid objects such as our cube.

While the fact that congenitally blind participants used different haptic strategies may be simply due to their higher training in using only the haptic modality, it is also possible that these differences could be partly due to divergent spatial strategies between congenitally blind, sighted and late blind persons. Previous studies suggested indeed that sighted individuals might prefer using an allocentric frame of reference (Noordzij et al., 2006; Pasqualotto et al., 2013) which, although accurate, may need more time to be built (Toroj and Szubielska, 2011). Even though we did not explicitly investigate this issue, two congenitally blind participants spontaneously reported they counted the number of pins of the cube faces to help memorizing pins configuration which suggests they were not using an allocentric strategy. This observation is also well in line with a previous finding showing that early blind subjects encoded 2D pattern elements by their location in a fixed coordinate system without visual representation (Vanlierde and Wanet-Defalque, 2004). Future studies might want to investigate in detail such cognitive aspects of haptic exploration using the iCube.

With our current data, it is difficult to conclude whether the difference between congenitally and late blind is due to the fact the former group has never experienced the visual world and, therefore, it has exploited the brain plasticity that strongly characterizes the early years of life (e.g., Kupers and Ptito, 2014) resulting in a stronger haptic ability (Theurel et al., 2013) or to the fact that haptic skills are simply more trained in the congenitally blind since they lived more "years of blindness."

Our congenitally blind group has experienced a mean of 35.5 years of blindness, whereas, this mean in the late blind group was 21.6 years. Future studies will be needed to compare exploration behavior of congenitally and late blind individuals having a similar amount of years of blindness (although, in this case, differing for age). On the other hand, we speculate that, since our late blind participants were probably fully blind for long enough to match the haptic expertise of the congenitally blind, the main difference between the two groups may lie in the extended haptic practice in the congenitally blind in their early years of life (Theurel et al., 2013; Amadeo et al., 2019).

One limitation of our study lies in the small sample size, particularly the congenitally blind group. This may have limited the possibility to spot other haptic differences between this group and late blind and sighted groups. However, specific differences between groups, that is, the mean number of active cells per sample and the variability in active cells across recording samples, were evidently large enough to be already detected with groups of such size. A second limitation lies in that information about Braille-reading ability in our blind participants was not available. There is evidence that experience in reading Braille is correlated with superior tactile acuity in passive tasks (Wong et al., 2011) and in tasks using Braille-like stimuli (e.g., Foulke and Warm, 1967; Grant et al., 2000). However, our task involved the active manipulation of a 3D object and the pins attached on its faces have different dimension (diameter: 3 mm; height: 2 mm) than Braille dots (diameter: 1.44 mm; height: 5 mm). More importantly, the spacing between pins in our configuration is in the order of centimeters whereas it is about 2.5 mm in the Braille. Therefore, our task did not involve any measure of tactile acuity at its limit of performance, as Wong et al. (2011) did. A third limitation is represented by the fact we used a cube-shaped object which imposes limits in the exploration behavior of participants and makes potentially difficult generalizing our results to objects with more complex shapes. Finally, subjects performed a small number of trials since we wanted to minimize the effort of participants. Therefore, we could not investigate in detail the temporal evolution of performance as well as possible changes in exploration strategies.

In conclusion, our study showed that congenitally, late blind and sighted participants did not differ in the haptic recognition accuracy of a three-dimensional object. However, we identified two exploratory strategies that differentiated congenitally blind from late blind and sighted individuals. The former group touched more cells simultaneously when exploring a face, suggesting that they could acquire more tactile information "at first glance." Furthermore, congenitally blind showed higher haptic velocity, that is, they changed more the pattern of touched cells from one recording sample to the next. Finally, we also found that the number of simultaneously touched cells negatively correlated with exploration duration suggesting that the ability to cover a larger tactile space while touching an object allows a more effective and faster exploration.

Future studies might want to verify whether we could use the sensorized cube to measure the haptic and spatial skills of different populations such as in the elderly. There is indeed evidence that cognitive decline may impair haptic object recognition (Kalisch et al., 2012) but the modulation of the exploratory procedures by age has not been investigated in detail yet.

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://doi.org/10.5281/zenodo.6539275.

Ethics statement

The studies involving human participants were reviewed and approved by Comitato Etico, ASL 3, Genova; Prot. IIT_UVIP_COMP_2019 N. 02/2020, 4 July 2020. 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

FL performed testing and data analysis. FL wrote this manuscript with contributions from MG and AS. All authors developed the study concept, contributed to the study design, and approved the final version of the manuscript for submission.

Funding

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (grant agreement nos. 948349, MYSpace and 804388, wHiSPER).

Acknowledgments

We thank Giulio Sandini, Antonio Maviglia, Diego Torazza, and Elio Massa for the ideation, design, and realization of the iCube, Marco Jacono for the development of the first version of the software, Marcello Goccia for the current version of the software running the iCube, and Alessia Tonelli and Alice Bollini for helping in organizing the tests with visually impaired persons. Special thanks to all participants.

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

References

Amadeo, M. B., Campus, C., and Gori, M. (2019). Impact of years of blindness on neural circuits underlying auditory spatial representation. *Neuroimage* 191, 140–149. doi: 10.1016/j.neuroimage.2019.01.073

Amedi, A., von Kriegstein, K., van Atteveldt, N. M., Beauchamp, M. S., and Naumer, M. J. (2005). Functional imaging of human crossmodal identification and object recognition. *Exp. Brain Res.* 166, 559–571. doi: 10.1007/s00221-005-2396-5

Benjamini, Y., and Hochberg, Y. (1995a). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57, 289–300. doi: 10.2307/2346101

Benjamini, Y., and Hochberg, Y. (1995b). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Bliss, I., Kujala, T., and Hämäläinen, H. (2004). Comparison of blind and sighted participants' performance in a letter recognition working memory task. *Cogn. Brain Res.* 18, 273–277. doi: 10.1016/J.COGBRAINRES.2003.10.012

Bodegård, A., Geyer, S., Grefkes, C., Zilles, K., and Roland, P. E. (2001). Hierarchical Processing of Tactile Shape in the Human Brain. *Neuron* 31, 317–328. doi: 10.1016/S0896-6273(01)00362-2

Burton, H. (2003). Visual cortex activity in early and late blind people. *J. Neurosci.* 23, 4005–4011. doi: 10.1523/JNEUROSCI.23-10-04005.2003

Campus, C., Sandini, G., Amadeo, M. B., and Gori, M. (2019). Stronger responses in the visual cortex of sighted compared to blind individuals during auditory space representation. *Sci. Rep.* 9:1935. doi: 10.1038/s41598-018-37821-y

Cattaneo, Z., and Vecchi, T. (2008). Supramodality Effects in Visual and Haptic Spatial Processes. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 631–642. doi: 10.1037/0278-7393.34.3.631

Cooke, T., Jäkel, F., Wallraven, C., and Bülthoff, H. H. (2007). Multimodal similarity and categorization of novel, three-dimensional objects. *Neuropsychologia* 45, 484–495. doi: 10.1016/j.neuropsychologia.2006.02.009

Davidson, P. W. (1972). Haptic judgments of curvature by blind and sighted humans. *J. Exp. Psychol.* 93, 43–55. doi: 10.1037/h0032632

Foulke, E., and Warm, J. S. (1967). Effects of Complexity and Redundancy on the Tactual Recognition of Metric Figures. *Percept. Mot. Skills* 25, 177–187. doi: 10.2466/pms.1967.25.1.177

Gaißert, N., Waterkamp, S., Fleming, R. W., and Bülthoff, I. (2012). Haptic Categorical Perception of Shape. *PLoS One* 7:e43062. doi: 10.1371/journal.pone.0043062

Giudice, N. A., Betty, M. R., and Loomis, J. M. (2011). Functional Equivalence of Spatial Images From Touch and Vision: Evidence From Spatial Updating in Blind and Sighted Individuals. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 621–634. doi: 10.1037/a0022331

Goldreich, D., and Kanics, I. M. (2003). Tactile Acuity is Enhanced in Blindness. *J. Neurosci.* 23, 3439–3445.

Gori, M., Sandini, G., Martinoli, C., and Burr, D. (2010). Poor haptic orientation discrimination in nonsighted children may reflect disruption of cross-sensory calibration. *Curr. Biol.* 20, 223–225. doi: 10.1016/j.cub.2009.11.069

Gori, M., Sandini, G., Martinoli, C., and Burr, D. C. (2014). Impairment of auditory spatial localization in congenitally blind human subjects. *Brain* 137, 288–293. doi: 10.1093/brain/awt311

Grant, A. C., Thiagarajah, M. C., and Sathian, K. (2000). Tactile perception in blind Braille readers: A psychophysical study of acuity and hyperacuity using gratings and dot patterns. *Percept. Psychophys.* 62, 301–312. doi: 10.3758/BF03205550

Heller, M. A. (1989). Picture and pattern perception in the sighted and the blind: The advantage of the late blind. *Perception* 18, 379–389. doi: 10.1068/p180379

Heller, M. A. (2002). Tactile picture perception in sighted and blind people. *Behav. Brain Res.* 135, 65–68. doi: 10.1016/S0166-4328(02)00156-0

James, T. W., Humphrey, G. K., Gati, J. S., Servos, P., Menon, R. S., and Goodale, M. A. (2002). Haptic study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia* 40, 1706–1714. doi: 10.1016/S0028-3932(02)00017-9

Kalisch, T., Kattenstroth, J.-C., Kowalewski, R., Tegenthoff, M., and Dinse, H. R. (2012). Cognitive and Tactile Factors Affecting Human Haptic Performance in Later Life. *PLoS One* 7:e30420. doi: 10.1371/journal.pone.0030420

Klatzky, R. L., and Lederman, S. J. (1992). Stages of manual exploration in haptic object identification. *Percept. Psychophys.* 52, 661–670. doi: 10.3758/BF03211702

Klatzky, R. L., Lederman, S. J., and Metzger, V. A. (1985). Identifying objects by touch: An "expert system.". *Percept. Psychophys.* 37, 299–302. doi: 10.3758/BF03211351

Kujala, T., Huotilainen, M., Sinkkonen, J., Ahonen, A. I, Alho, K., Hämäläinen, M. S., et al. (1995). Visual cortex activation in blind humans during sound discrimination. *Neurosci. Lett.* 183, 143–146. doi: 10.1016/0304-3940(94)11135-6

Kupers, R., and Ptito, M. (2014). Compensatory plasticity and cross-modal reorganization following early visual deprivation. *Neurosci. Biobehav. Rev.* 41, 36–52. doi: 10.1016/j.neubiorev.2013.08.001

Lacey, S., Peters, A., and Sathian, K. (2007). Cross-Modal Object Recognition Is Viewpoint-Independent. *PLoS One* 2:e890. doi: 10.1371/journal.pone.0000890

Lacey, S., and Sathian, K. (2011). "Chapter 11 - Multisensory object representation: Insights from studies of vision and touch," in *Enhancing Performance for Action and Perception*, eds A. M. Green, C. E. Chapman, J. F. Kalaska, and F. Lepore (Amsterdam: Elsevier), 165–176. doi: 10.1016/B978-0-444-53752-2.00006-0

Lacey, S., and Sathian, K. (2014). Visuo-haptic multisensory object recognition, categorization, and representation. *Front. Psychol.* 5:730. doi: 10.3389/fpsyg.2014.00730

Lawton, C. A. (1994). Gender differences in way-finding strategies: Relationship to spatial ability and spatial anxiety. *Sex Roles* 30, 765–779. doi: 10.1007/BF01544230

Lederman, S. J., and Klatzky, R. L. (1987). Hand Movements: A window into haptic object recognition. *Cogn. Psychol.* 19, 342–368.

Lederman, S. J., and Klatzky, R. L. (1993). Extracting object properties through haptic exploration. *Acta Psychol.* 84, 29–40. doi: 10.1016/0001-6918(93)90070-8

Lederman, S. J., Klatzky, R. L., Chataway, C., and Summers, C. D. (1990). Visual mediation and the haptic recognition of two-dimensional pictures of common objects. *Percept. Psychophys.* 47, 54–64. doi: 10.3758/BF03208164

Leo, F., Cocchi, E., Ferrari, E., and Brayda, L. (2020). "Maps as Ability Amplifiers: Using Graphical Tactile Displays to Enhance Spatial Skills in People Who Are Visually Impaired," in *Haptic Interfaces for Accessibility, Health, and Enhanced Quality of Life*, eds T. McDaniel and S. Panchanathan (Berlin: Springer), 65–88. doi: 10.1007/978-3-030-34230-2_3

Leo, F., Sandini, G., and Sciutti, A. (2022). Mental Rotation Skill Shapes Haptic Exploration Strategies. *IEEE Trans. Haptics* 15, 339–350. doi: 10.1109/TOH.2022.3162321

Leo, F., Tinti, C., Chiesa, S., Cavaglià, R., Schmidt, S., Cocchi, E., et al. (2018). Improving spatial working memory in blind and sighted youngsters using programmable tactile displays. *SAGE Open Med.* 6:205031211882002. doi: 10.1177/2050312118820028

Loomis, J. M., Klatzky, R. L., and Giudice, N. A. (2013). "Representing 3D space in working memory: Spatial images from vision, hearing, touch, and language," in *Multisensory imagery*, eds S. Lacey and R. Lawson (Berlin: Springer), 131–155. doi: 10.1007/978-1-4614-5879-1_8

Meneghetti, C., Pazzaglia, F., and De Beni, R. (2011). Spatial mental representations derived from survey and route descriptions: When individuals prefer extrinsic frame of reference. *Learn. Individ. Differ.* 21, 150–157. doi: 10.1016/j.lindif.2010.12.003

Morey, R. D., and Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychol. Methods* 16, 406–419. doi: 10.1037/a0024377

Morrongiello, B. A., Humphrey, G. K., Timney, B., Choi, J., and Rocca, P. T. (1994). Tactual object exploration and recognition in blind and sighted children. *Perception* 23, 833–848. doi: 10.1068/p230833

Newell, F. N., Ernst, M. O., Tjan, B. S., and Bülthoff, H. H. (2001). Viewpoint Dependence in Visual and Haptic Object Recognition. *Psychol. Sci.* 12, 37–42. doi: 10.1111/1467-9280.00307

Noordzij, M. L., Zuidhoek, S., and Postma, A. (2006). The influence of visual experience on the ability to form spatial mental models based on route and survey descriptions. *Cognition* 100, 321–342. doi: 10.1016/j.cognition.2005.05.006

Norman, J. F., and Bartholomew, A. N. (2011). Blindness enhances tactile acuity and haptic 3-D shape discrimination. *Atten. Percept. Psychophys.* 73, 2323–2331. doi: 10.3758/s13414-011-0160-4

Norman, J. F., Norman, H. F., Clayton, A. M., Lianekhammy, J., and Zielke, G. (2004). The visual and haptic perception of natural object shape. *Percept. Psychophys.* 66, 342–351. doi: 10.3758/BF03194883

Occelli, V., Lacey, S., Stephens, C., John, T., and Sathian, K. (2016). Haptic Object Recognition is View-Independent in Early Blind but not Sighted People. *Perception* 45, 337–345. doi: 10.1177/0301006615614489

Ottink, L., Hoogendonk, M., Doeller, C. F., Van der Geest, T. M., and Van Wezel, R. J. A. (2021). Cognitive map formation through haptic and visual exploration of tactile city-like maps. *Sci. Rep.* 11:15254. doi: 10.1038/s41598-021-94778-1

Pasqualotto, A., Spiller, M. J., Jansari, A. S., and Proulx, M. J. (2013). Visual experience facilitates allocentric spatial representation. *Behav. Brain Res.* 236, 175–179. doi: 10.1016/j.bbr.2012.08.042

Penfield, W., and Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain J. Neurol.* 60, 389.

Pensky, A. E. C., Johnson, K. A., Haag, S., and Homa, D. (2008). Delayed memory for visual—haptic exploration of familiar objects. *Psychon. Bull. Rev.* 15, 574–580. doi: 10.3758/PBR.15.3.574

Picard, D., and Lebaz, S. (2012). Identifying Raised-Line Drawings by Touch: A Hard but Not Impossible Task. *J. Vis. Impair. Blind.* 106, 427–431.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *J. Exp. Psychol. Hum. Learn. Mem.* 2, 509–522. doi: 10.1037/0278-7393.2.5.509

Puce, A., Allison, T., Gore, J. C., and McCarthy, G. (1995). Face-sensitive regions in human extrastriate cortex studied by functional MRI. *J. Neurophysiol.* 74, 1192–1199. doi: 10.1152/jn.1995.74.3.1192

Sathian, K., and Stilla, R. (2010). Cross-modal plasticity of tactile perception in blindness. *Restor. Neurol. Neurosci.* 28, 271–281. doi: 10.3233/RNN-2010-0534

Sciutti, A., Damonte, F., Alloisio, M., and Sandini, G. (2019). Visuo-Haptic Exploration for Multimodal Memory. *Front. Integr. Neurosci.* 13:15. doi: 10.3389/fnint.2019.00015

Sciutti, A., and Sandini, G. (2019). "The role of object motion in visuo-haptic exploration during development," in *2019 Joint IEEE 9th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, (Oslo: IEEE), doi: 10.1109/DEVLRN.2019.8850687

Sciutti, A., and Sandini, G. (2020). To move or not to move: Development of fine-tuning of object motion in haptic exploration. *IEEE Trans. Cogn. Dev. Syst.* 14, 366–374. doi: 10.1109/TCDS.2020.3034014

Szubielska, M., and Zabielska-Mendyk, E. (2018). Mental rotation of figures explored by touch: A study of congenitally blind and sighted individuals. *Ann. Psychol.* 21, 35–51. doi: 10.18290/rpsych.2018.21.1-3

Taylor, H. A., and Tversky, B. (1992). Spatial mental models derived from survey and route descriptions. *J. Mem. Lang.* 31, 261–292. doi: 10.1016/0749-596X(92)90014-O

Theurel, A., Witt, A., Claudet, P., Hatwell, Y., and Gentaz, E. (2013). Tactile picture recognition by early blind children: The effect of illustration technique. *J. Exp. Psychol. Appl.* 19, 233–240. doi: 10.1037/a0034255

Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature* 381, 520–522. doi: 10.1038/381520a0

Toroj, M., and Szubielska, M. (2011). Prior visual experience, and perception and memory of shape in people with total blindness. *Br. J. Vis. Impair.* 29, 60–81. doi: 10.1177/0264619610387554

Vanlierde, A., and Wanet-Defalque, M.-C. (2004). Abilities and strategies of blind and sighted subjects in visuo-spatial imagery. *Acta Psychol.* 116, 205–222. doi: 10.1016/j.actpsy.2004.03.001

Vecchi, T., Tinti, C., and Cornoldi, C. (2004). Spatial memory and integration processes in congenital blindness. *Neuroreport* 15, 2787–2790.

Venables, W. N., and Ripley, B. D. (2002). *Modern Applied Statistics with S. Fourth Edition*. Berlin: Springer.

Vinter, A., Fernandes, V., Orlandi, O., and Morgan, P. (2012). Exploratory procedures of tactile images in visually impaired and blindfolded sighted children: How they relate to their consequent performance in drawing. *Res. Dev. Disabil.* 33, 1819–1831. doi: 10.1016/j.ridd.2012.05.001

Vinter, A., Orlandi, O., and Morgan, P. (2020). Identification of Textured Tactile Pictures in Visually Impaired and Blindfolded Sighted Children. *Front. Psychol.* 11:345. doi: 10.3389/fpsyg.2020.00345

Wallraven, C., and Dopjans, L. (2013). Visual experience is necessary for efficient haptic face recognition. *Neuroreport* 24, 254–258. doi: 10.1097/WNR.0b013e32835f00c0

Wan, C. Y., Wood, A. G., Reutens, D. C., and Wilson, S. J. (2010). Congenital blindness leads to enhanced vibrotactile perception. *Neuropsychologia* 48, 631–635. doi: 10.1016/j.neuropsychologia.2009.10.001

Wijntjes, M. W. A., van Lienen, T., Verstijnen, I. M., and Kappers, A. M. L. (2008). Look what I have felt: Unidentified haptic line drawings are identified after sketching. *Acta Psychol.* 128, 255–263. doi: 10.1016/j.actpsy.2008.01.006

Withagen, A., Kappers, A. M. L., Vervloed, M. P. J., Knoors, H., and Verhoeven, L. (2012). Haptic object matching by blind and sighted adults and children. *Acta Psychol.* 139, 261–271. doi: 10.1016/j.actpsy.2011.11.012

Wong, M., Gnanakumaran, V., and Goldreich, D. (2011). Tactile spatial acuity enhancement in blindness: Evidence for experience-dependent mechanisms. *J. Neurosci.* 31, 7028–7037. doi: 10.1523/JNEUROSCI.6461-10.2011

Yue, X., Tjan, B. S., and Biederman, I. (2006). What makes faces special? *Vision Res.* 46, 3802–3811. doi: 10.1016/j.visres.2006.06.017

frontiers | Frontiers in Human Neuroscience

TYPE Brief Research Report
PUBLISHED 13 October 2022
DOI 10.3389/fnhum.2022.1026056

OPEN ACCESS

EDITED BY
Alessia Tonelli,
Italian Institute of Technology
(IIT), Italy

REVIEWED BY
Andrew Joseph Kolarik,
University of East Anglia,
United Kingdom
Erol J. Ozmeral,
University of South Florida,
United States

*CORRESPONDENCE
Elena Gessa
elena.gessa@unitn.it
Chiara Valzolgher
chiara.valzolgher@unitn.it

SPECIALTY SECTION
This article was submitted to
Sensory Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 23 August 2022
ACCEPTED 21 September 2022
PUBLISHED 13 October 2022

CITATION
Gessa E, Giovanelli E, Spinella D,
Verdelet G, Farnè A, Frau GN, Pavani F
and Valzolgher C (2022) Spontaneous
head-movements improve sound
localization in aging adults with
hearing loss.
Front. Hum. Neurosci. 16:1026056.
doi: 10.3389/fnhum.2022.1026056

Spontaneous head-movements improve sound localization in aging adults with hearing loss

Elena Gessa[1]*, Elena Giovanelli[1], Domenico Spinella[2], Grégoire Verdelet[3,4], Alessandro Farnè[1,3,4], Giuseppe Nicolò Frau[2], Francesco Pavani[1,3] and Chiara Valzolgher[1,3]*

[1]Center for Mind/Brian Sciences - CIMeC, University of Trento, Rovereto, Italy, [2]"S. Maria del Carmine" Hospital, Rovereto, Italy, [3]Integrative, Multisensory, Perception, Action and Cognition Team-IMPACT, Centre de Recherche en Neuroscience de Lyon, University Lyon 1, Lyon, France, [4]Neuro-immersion, Centre de Recherche en Neuroscience de Lyon, Lyon, France

Moving the head while a sound is playing improves its localization in human listeners, in children and adults, with or without hearing problems. It remains to be ascertained if this benefit can also extend to aging adults with hearing-loss, a population in which spatial hearing difficulties are often documented and intervention solutions are scant. Here we examined performance of elderly adults (61–82 years old) with symmetrical or asymmetrical age-related hearing-loss, while they localized sounds with their head fixed or free to move. Using motion-tracking in combination with free-field sound delivery in visual virtual reality, we tested participants in two auditory spatial tasks: front-back discrimination and 3D sound localization in front space. Front-back discrimination was easier for participants with symmetrical compared to asymmetrical hearing-loss, yet both groups reduced their front-back errors when head-movements were allowed. In 3D sound localization, free head-movements reduced errors in the horizontal dimension and in a composite measure that computed errors in 3D space. Errors in 3D space improved for participants with asymmetrical hearing-impairment when the head was free to move. These preliminary findings extend to aging adults with hearing-loss the literature on the advantage of head-movements on sound localization, and suggest that the disparity of auditory cues at the two ears can modulate this benefit. These results point to the possibility of taking advantage of self-regulation strategies and active behavior when promoting spatial hearing skills.

KEYWORDS
aging, sound localization, head movements, hearing loss, virtual reality

Introduction

Age-related hearing loss (ARHL) is a major issue for individuals and the society. It develops gradually, often in a subtle fashion: at first it reduces the detection of high-pitched sounds and speech comprehension in noisy environments, then it progresses into a more generalized difficulty in understanding conversations

(Davis, 2019). Although the impact of ARHL has been primarily investigated in relation to speech comprehension (Noble et al., 1995), it impacts auditory environment more broadly. In particular, it effects the ability to localize sounds in space, by changing binaural and monaural auditory cues available at each ear. Binaural cues (interaural level difference, ILD and interaural time difference, ITD) play a key role when determining the horizontal direction of sounds. In normal-hearing, sound localization exploits primarily low-frequency (<1400 Hz) ITD cues, with a secondary role for high-frequency (>4000 Hz) ITD and ILD cues (see Macpherson and Middlebrooks, 2002). Monaural cues depend upon the direction of sounds with respect to the head and the external ear, and contribute primarily to front-back disambiguation, elevation estimation, and distance perception of sounds. They can be successfully extracted only from high-frequency sounds (>4000 Hz Middlebrooks, 2015). The high-frequencies loss of ARHL impacts high-frequency ITD and ILD cues, as well as monaural cues. Moreover, with declines in neural synchrony and reduced central inhibition related to advancing age, processing of auditory cues is hindered even more (Eddins et al., 2018). This results in worse front-back discrimination and impoverished localization of sounds on the vertical plane in elderly adults (Rakerd et al., 1998). Finally, in the case of asymmetric ARHL the imbalance between binaural cues could result in poor localization performances on the horizontal plane, considering the lack of high-frequency monaural cues that normally compensate for the inability of extracting binaural cues (Kumpik and King, 2019). In this context of reduced peripheral cues at the ears, how can aging adults improve their sound localization skills?

Head-movement during sound is a spontaneous and ubiquitous behavior that impacts on sound localization. Head-movements change the available auditory cues: rotations around the vertical axis modify ITD and ILD cues, whereas tilting the head impacts on monaural cues (Perrett and Noble, 1997a; Kato et al., 2003). Head-movements present several orientations of the ears to the sound and therefore provide richer and more dynamic auditory cues (Pollack and Rose, 1967). Although the role of head movements in spatial hearing has been advocated since the first half of the last century (Wallach, 1940), systematic investigations have only started in the last decades, also as a consequence of greater availability of motion-tracking technologies. Head-movements during sound improve sound localization in normal-hearing adults on both horizontal and vertical dimensions (Perrett and Noble, 1997a) and reduce front-back discrimination errors (Iwaya et al., 2003). In addition, head-movements improve sound localization in hearing-impaired adults (Brimijoin et al., 2012) and cochlear implant users (adults: Pastore et al., 2018; children: Coudert et al., 2022). If head-movements can improve sound localization in the context of ARHL remains, to the best of our knowledge, an open question.

This study aimed to examine if head-movements improve sound localization in aging adults with ARHL. Previous studies asked participants to perform stereotyped movements (e.g., keep their movements slow, continuous and in a ±30° range; Pastore et al., 2018) or forced passive head-movement through robotic control of the participant's head (Thurlow and Runge, 1967). Here, we opted for inviting participants to produce spontaneous head-movements while the sound was playing, without giving any specific instructions as to movement speed or extension (as in Coudert et al., 2022). To measure sound localization and head-movements, we exploited a visual virtual reality and motion tracking approach (Valzolgher et al., 2020a; Coudert et al., 2022), which allows extensive control over the audio-visual stimulation delivered to participants. We asked participants to localize sounds in a visual virtual reality scenario while recording their head-movements in real-time under two listening conditions: head-fixed and head free to move. Each participant performed two auditory spatial tasks: front-back discrimination and 3D sound localization in front space (participants responded using a hand-held tool and we measured their responses in azimuth, elevation and distance). We enrolled aged participants with different degrees of hearing-impairment, who were divided in two groups differing for hearing asymmetry: symmetrical and asymmetrical ARHL.

We expected spontaneous head-movements to facilitate sound localization in both tasks. Head-rotations around the vertical axis modify time of arrival and level of sounds at the two ears and therefore each degree of rotation is associated with different auditory binaural cues. Dynamical changes of binaural cues enable a more reliable selection between different possible sound sources that vary in 3D space (McAnally and Martin, 2014). Regarding front-back discrimination, rotations along the head vertical axis transform front-back confusion into left-right discrimination, increasing the possibility of using binaural cues. We expected a benefit in sound localization when head is free to move, particularly for asymmetrical ARHL participants who are more likely to experience auditory cues ambiguities during head-fixed listening.

Materials and methods

Participants

Sixteen participants (mean age 71, SD = 6.51, range = [61–82], 7 males) took part in the study. Sample size was driven by previous studies that investigated head-movements effects on sound localization in normal-hearing (Perrett and Noble, 1997b: $N = 16$) and hearing-impaired participants (Pastore et al., 2018: 7 listeners bilaterally implanted with cochlear implants, 5 of the patients with one implant turned off). Half of participants suffered symmetrical hearing-loss, with an average of 6.88 dB HL (SD = 3.23, range = [3–13]) difference in hearing threshold

TABLE 1 Personal and audiometric characteristics of participants.

Participant	Group	Age	Sex	Worst ear	Threshold best ear	Threshold worst ear	Disparity between the two ears
1	Asymmetric	68	M	Left	15 dB HL (normal)	63 dB HL (moderate severe)	48 dB HL
2	Asymmetric	69	M	Left	21 dB HL (slight)	58 dB HL (moderate)	37 dB HL
3	Asymmetric	77	M	Right	16 dB HL (slight)	58 dB HL (moderate severe)	42 dB HL
4	Asymmetric	71	F	Right	23 dB HL (slight)	51 dB HL (moderate)	28 dB HL
5	Asymmetric	81	F	Left	38 dB HL (mild)	63 dB HL (moderate severe)	25 dB HL
6	Asymmetric	62	M	Left	23 dB HL (slight)	60 dB HL (moderate severe)	37 dB HL
7	Asymmetric	63	M	Right	18 dB HL (slight)	78 dB HL (severe)	60 dB HL
8	Asymmetric	73	F	Left	18 dB HL (slight)	56 dB HL (moderate severe)	38 dB HL
9	Symmetric	65	F	Left	11 dB HL (normal)	15 dB HL (normal)[1]	4 dB HL
10	Symmetric	77	M	Left	29 dB HL (mild)	37 dB HL (mild)	8 dB HL
11	Symmetric	69	F	Left	22 dB HL (slight)	30 dB HL (mild)	8 dB HL
12	Symmetric	61	F	Right	18 dB HL (slight)	25 dB HL (slight)	7 dB HL
13	Symmetric	69	F	Left	16 dB HL (slight)	29 dB HL (mild)	13 dB HL
14	Symmetric	77	F	Left	33 dB HL (mild)	37 dB HL (mild)	4 dB HL
15	Symmetric	82	M	Left	41 dB HL (moderate)	44 dB HL (moderate)	3 dB HL
16	Symmetric	72	F	Right	28 dB HL (mild)	36 dB HL (mild)	8 dB HL

Personal and audiometric characteristics of participants. Classification criteria refer to Clark (1981). Uses and abuses of hearing loss classification. Asha, 23(7), 493–500. For each ear we tested frequencies at 125, 250, 500, 750, 1000, 1500, 2000, 3000, 4000, 6000, and 8000 Hz, through a pure tone audiometry.

between the two ears. Hearing thresholds were 31.63 dB HL (SD = 8.94, range = [15–44]) in the worse ear, and 24.75 dB HL (SD = 9.85, range = [11–41]) in the best ear. The remaining half of participants suffered from asymmetrical hearing loss, with a difference in hearing threshold between ears of 39.38 dB HL (SD = 11.06, range = [25–60]). Hearing thresholds were 60.88 dB HL (SD = 7.94, range = [51–78]) in the worse ear, and 21.50 dB HL (SD = 7.31, range = [15–38]) in the best ear. All had normal or correct-to-normal vision. See Table 1 for further details.

Asymmetrical ARHL participants were recruited at the otolaryngology department of "S. Maria del Carmine" hospital in Rovereto (Italy), symmetrical ARHL participants were recruited through advertisement. All volunteers gave their informed consent before starting the experiment, which was approved by the Ethics Committee of the University of Trento (protocol number: 2019-037). The inclusion concerned only individuals without hearing aids, who did not use drugs, or reported a history of neurological or psychiatric problems. All participants completed the Montreal Cognitive Assessment test (MoCA, Italian version: Conti et al., 2015) to exclude possible cognitive decline and all obtained normal scores for their age.

Stimuli

The auditory target was a white-noise (43–22000 Hz; sample rate: 44100 Hz), with an 80% amplitude-modulation at 2.5 Hz. We adopted this broadband stimulus to preserve processing of all frequencies available to each ear (Hofman et al., 1998; Savel et al., 2009; Gaveau et al., 2022; Valzolgher et al., 2022a). Moreover, we modulate noise's amplitude to facilitate ITD processing by reducing phase ambiguities (Macpherson and Middlebrooks, 2002). Sound was delivered at about 75 dB SPL, as measured from the participant's head using a decibel meter (TES1350A). Each auditory target lasted 5 seconds, to allow participants enough time to make spontaneous head-movements during the head-free condition. Auditory targets were delivered at pre-determined positions in each trial (see Procedure).

Apparatus

The experiment was run using the HTC Vive system, a virtual reality and motion tracking device [see Valzolgher et al., 2020a]. This system comprised one Head-Mounted Display (HMD, resolution: 1080 × 1200 px, Field of View (FOV): 110°, Refresh rate: 90 Hz) for the presentation of visual stimuli; one hand-held tracker used by participants to collect pointing responses; one tracker placed above the speaker to monitor its position in real-time; one hand-held controller used by the experimenter to record the responses. Finally, two lighthouse base stations scanned the position of the HMD, the trackers and the controller in real-time. The HTC Vive system and the lighthouse base stations were controlled by a LDLC ZALMAN computer (OS: Windows 10 (10.0.0) 64bit; Graphic card: NVIDIA GeForce GTX 1060 6 GB; Processor: Intel Core i7-770 K, Quad-Core 4.2/4.5 GHz Turbo, Cache 8 Mo, TDP 95 W), using the Steam VR software and the development platform Unity.

All target sounds were delivered using a single real speaker (JBL GO Portable, 68.3 × 82.7 × 30.8 mm, Output Power 3.0 W, Frequency response 180–20 kHz, Signal-to-noise ratio >80 dB) at pre-determined positions within reaching space (see Procedure).

Procedure

After a preliminary description of the experiment, participants sat on a rotating chair in the center of a room (3 × 4 × 5 m) and wore the HMD. They were immersed in a virtual empty room, with the exact same metrics as the real room in the laboratory. The rationale for showing a visual virtual environment was twofold: first, it allowed to eliminate all visual cues about the sound source position (i.e., the loudspeaker); second, it allowed to present participants with a visual environment that was devoid of furniture. A visible room is a more ecological compared to a fully dark environment during sound localization because it can provide visual references that guide acoustic space perception [Majdak et al., 2010; see also Valzolgher et al., 2020a]. In this virtual scenario, participants saw the tracker that they held in their hand, to allow more accurate pointing movements to the perceived sound source. Virtual reality has been used previously with elderly participants, with no harmful outcomes or stressful situations reported (Crespo et al., 2016). Likewise, no participant tested in the present work reported motion-sickness or discomfort.

Our experimental setup allowed to deliver sounds at pre-determined positions defined in head-centered coordinates at the beginning of each trial. Specifically, the system computed the pre-determined position in 3D space with respect to the center of the head and the interaural axis, and gave the experimenter visual cues (on a dedicated monitor) to guide the loudspeaker to the exact target position with a 5 cm tolerance (see Gaveau et al., 2022). The experimenter held the speaker in the target position for sound emission with her hand. Crucial to this procedure was the calibration of head position, which occurred each time the HMD was worn. This calibration was performed by marking with the experimenter's hand-held controller the position of the left and right ears of the participant.

After the participant familiarized with the virtual environment, two sound localization tasks were performed: front-back discrimination (see Front-back discrimination task) and 3D sound localization in front space (see 3D sound localization in front space task). Participants completed each task under two listening conditions: head-fixed or head free to move. During the head free condition, no suggestions were given regarding how to move the head and in this sense, movements represent a spontaneous strategy. The order of tasks and listening conditions were counterbalanced across participants. The entire procedure took 2 h on average, including preparation and pauses, with a VR immersion of 75 min.

Participants received instructions in the HMD to acquire a front facing posture at the beginning of each trial. Specifically, they saw in the HMD their head direction (indicated by a blue cross) and were instructed to align it with a white cross in the center of the virtual room. As soon as the two crosses were aligned, and the experimenter brought the loudspeaker to the pre-determined target position, the crosses disappeared and the sound was delivered. This approach allows participants to achieve a replicable sound target locations across trials, without using an external constraint (e.g., a chin-rest) which would have been incompatible with the free-head movement condition.

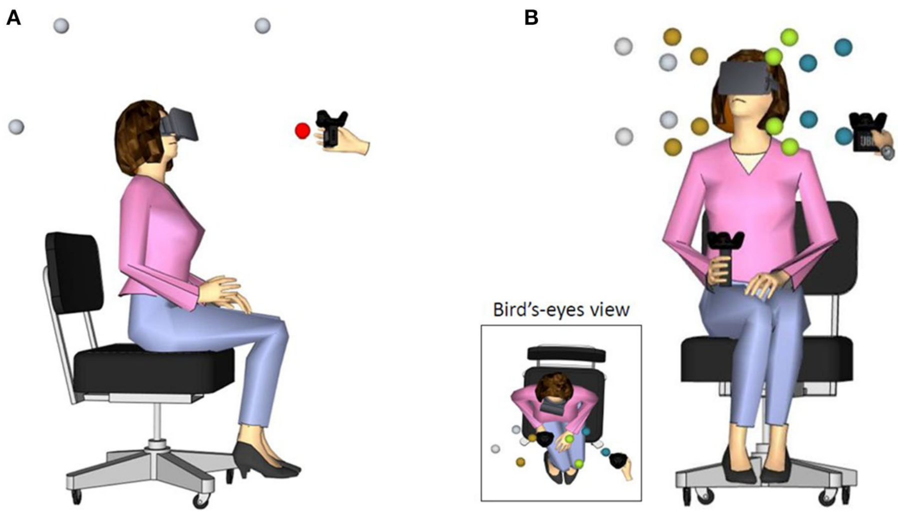


FIGURE 1
Schematic description of participant wearing the HMD in **(A)** the front-back discrimination task (lateral view) and **(B)** in the 3D sound localization in front space task (front and bird's-eye view). The spheres around the participant's head indicate the target positions. Different azimuth positions are marked by different colors only for illustrative purposes. The experimenter brought the tracked speaker (also shown in figure) at the pre-determined location identified in each trial.

Front-back discrimination task

In the front-back discrimination task, target positions were arranged along the participant's mid-sagittal plane, two at the front and two at the back (see Figure 1A). Specifically, the tracker connected to the loudspeaker was placed at 0° and 45° in front space, and at 135° and 180° in back space. All targets were delivered at about 50 centimeters from the center of the participant's head. Participants were instructed to listen to the sound and wait until its end before responding. They had to report verbally if the sound was emitted from front or back space. Responses were saved by the experimenter through the hand-held controller. No performance feedback was provided. Ten practice trials were included at the beginning of each block, to allow familiarization with the procedure. A total of 32 trials (8 repetitions for each of the 4 positions) were presented in randomized order within each listening condition.

3D sound localization in front space task

In the 3D sound localization task, target positions were in front space, at four azimuth (± 40° and ± 20° with respect to the midsagittal plane), two elevation (– 25° and 15° with respect to the plane passing through the ears) and two distances (35 or 55 centimeters from the center of the participant's head; see Figure 1B). Participants listened to the sound and wait until its end before responding. During sound emission, they kept their right hand holding the tracker stationary at the chest. When the sound ended, they were instructed to move the controller to the perceived position of the sound and validate their response by pressing a button on the tracker. Then, the experimenter triggered the beginning of the following trial. No performance feedback was provided [for a similar procedure see also Valzolgher et al., 2020a]. Ten practice trials were included at the beginning of each block, to allow familiarization with the procedure. A total of 48 trials (4 repetitions for each of the 16 positions) were presented in randomized order within each listening condition.

Analysis

All data were analyzed using Linear Mixed Effects (LME) or Generalized Linear Mixed Effects (GLME) models in R studio with the packages lme4 (Bates et al., 2014), car (Fox and Weisberg, 2020), and lmerTest (Kuznetsova et al., 2017). When appropriate, we corrected the skewness of distributions by log-transforming the variables. The raw data can be retrieved from osf.io/57chk. Details of kinematic analyses could be found in Supplementary Results.

To analyze the performance, we measured error rates for front-back discrimination task and average 3D errors for sound localization in front space. 3D errors represent the distance in centimeters between perceived positions of sources and the actual speaker's location. We then analyzed the average error on azimuth (in degrees), elevation (in degrees), and distance (in centimeters).

Results

As instructed, participants refrained from moving the head in the head-fixed condition and made spontaneous head-movements in the head-free condition. Occasional trials

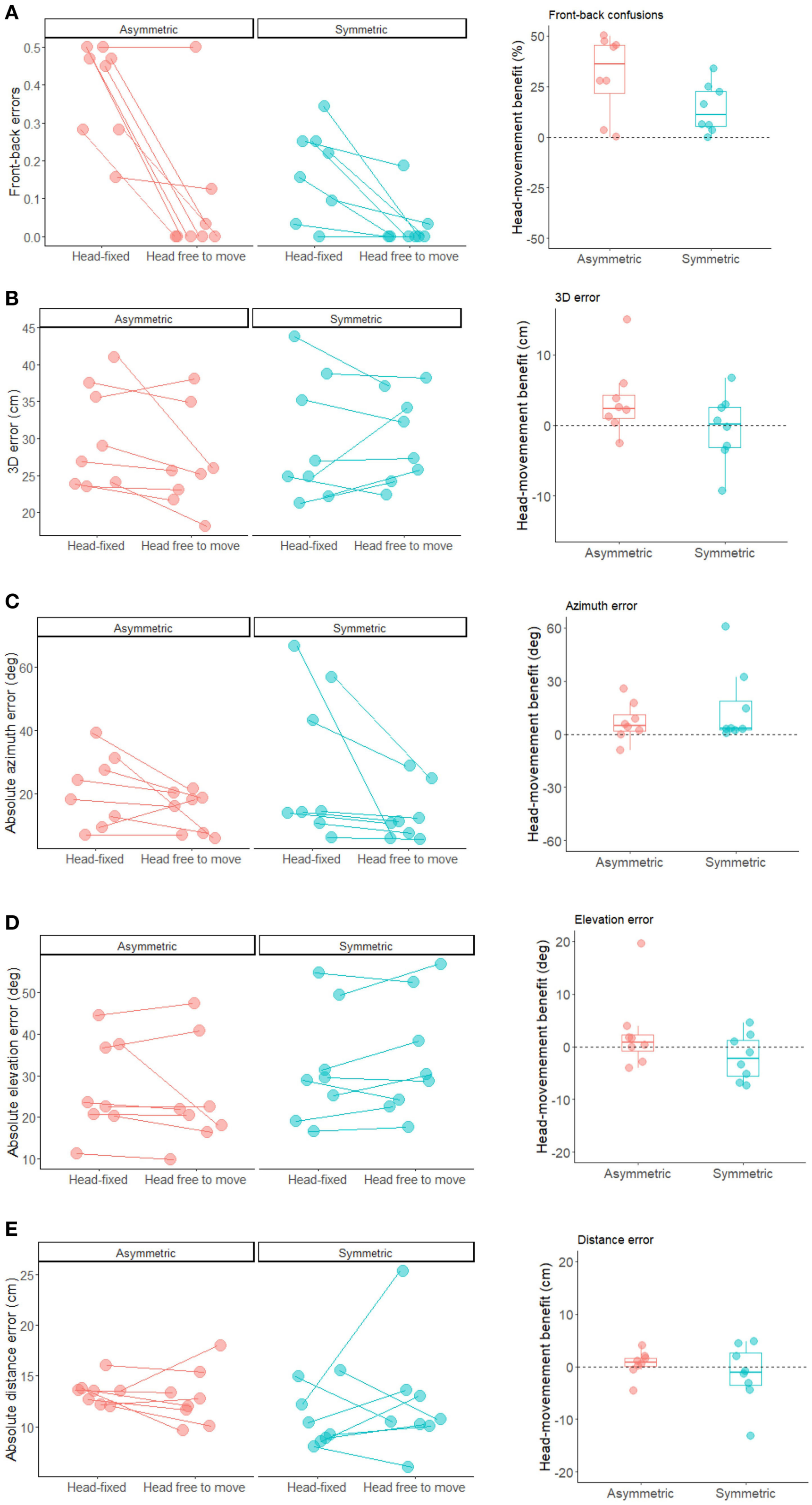


FIGURE 2
Individual performance (left plots) and head-movement benefit (right plots) for each participant in the front-back task **(A)** as well as the 3D sound localization task **(B)** 3D error; **(C)** azimuth error; **(D)** elevation error; **(E)** distance error. The head-movement benefit was calculated as the difference between errors in the head-free vs. head-fixed listening condition. Positive values indicate better performance when the head was free to move.

TABLE 2 Effects of spontaneous head-movements on symmetrical and asymmetrical ARHL participants.

	Symmetrical		Asymmetrical	
	Head free to move	**Head static**	**Head free to move**	**Head static**
3D sound localization in front space				
Azimuth	13.38 (± 8.79)	28.24 (± 23.70)	14.39 (± 6.60)	21.19 (± 11.35)
Elevation	33.85 (± 14.16)	31.81 (± 13.51)	24.65 (± 12.72)	27.12 (± 11.13)
Distance	12.43 (± 6.67)	10.94 (± 2.94)	12.85 (± 2.78)	13.41 (± 1.25)
3D error	30.15 (± 6.91)	29.73 (± 8.37)	26.58 (± 6.68)	30.15 (± 6.91)
Front-back discrimination				
Error rate	0.03 (± 0.07)	0.17 (± 0.12)	0.08 (± 0.17)	0.39 (± 0.13)

Effects of spontaneous head-movements on symmetrical and asymmetrical ARHL participants. Means with standard deviation in parenthesis. Note that azimuth and elevation errors are expressed in degrees, distance and 3D error in cm.

with head-movements in the head-fixed condition were removed from the analyses (0.29% of trials in the front-back discrimination task; 1.37% of trials in the 3D sound localization task). In the front-back discrimination task, during the head-free condition, participants made 4.0 (SD = 1.9) spontaneous head-movements, with a horizontal head-rotation extent of 46.8° and a vertical head-rotation extent of 18.6°. Instead, in the 3D localization task, they made 3.3 (SD = 1.5) spontaneous head-movements, with a horizontal head-rotation extent of 37.8° and a vertical head-rotation extent of 24.2°.

To study the effect of listening condition on front-back discrimination performance, we entered the binomial responses of each participant in a GLME model (family = binomial), using listening condition and group as categorical fixed effects and the participants' intercept as a random effect. Percent errors in front-back discrimination were smaller for symmetrical (9.7% ± 10%) compared to asymmetrical hearing-loss participants (23.5% ± 11.8%; main effect of group: $X^2(1) = 9.11$, $p = 0.003$). Importantly, spontaneous head-movements reduced percent errors (5.5% ± 13.0%) compared to the head-fixed condition (27.8% ± 16.6%) for both groups (main effect of listening condition: $X^2(1) = 58.98$, $p < 0.001$; see Figure 2A).

To study the 3D sound localization in front plane task, we computed the distance in centimeters between the 3D position of the sound source indicated in each trial and the actual 3D location of the speaker (i.e., 3D error). Trials in which participants moved the controller held in their hand during sound emission were rejected (2.0%). Additionally, 12.4% deviant data-points were excluded from the analyses following quantile-to-quantile plot inspection. We entered the 3D error in LME model, using listening condition and group as categorical fixed effects and the participants' intercept as a random effect. Spontaneous head-movements reduced the 3D error for participants with asymmetrical hearing-loss, whereas this benefit was not evident in participants with symmetrical hearing-loss resulting in a significant two-way interaction ($X^2(1) = 16.32$, $p < 0.001$; see Figure 2B and Table 2). The main effect of listening condition also reached significance ($X^2(1) = 20.63$, $p < 0.001$), but subsidiary to the higher order interaction.

We also examined the effect of listening posture on absolute localization errors in each dimension separately (i.e., azimuth, elevation and distance; see Table 2), using a LME models similar to the one adopted for the 3D error. Quantile-to-quantile plot inspection led to exclusion of 7.9% for azimuth, 6.4% for elevation, and 6.1% for distance. For azimuth (see Figure 2C), we found a main effect of listening condition ($X^2(1) = 32.92$, $p < 0.001$), caused by smaller absolute errors in the head-free (13.8° ± 7.9°) compared to the head-fixed condition (25.6° ± 21.2°). For elevation, we also found a main effect of listening condition ($X^2(1) = 5.32$, $p < 0.02$) and a two-way interaction ($X^2(1) = 7.16$, $p = 0.007$), caused by larger benefits of head-movements for participants with asymmetrical than symmetrical hearing-loss (Figure 2D). Finally, for distance, no main effect or interaction emerged (all p-values > 0.21; Figure 2E).

While the main purpose of our experiment was to examine effects of head-movements on sound localization performance, in Supplementary Results we also report our analyses on head-movements during the head free to move condition.

Discussion

In the present study, we examined if spontaneous head-movements can improve sound localization in aging adults with symmetrical or asymmetrical ARHL. We examined the ability to discriminate between sounds presented from front and back space and the ability to localize 3D sounds in front space under two different listening conditions: head-fixed and head free to move during sound emission.

Our findings show that spontaneous head-movements during sound presentation reduce front-back error and facilitate 3D sound localization in front space. The latter effect was more consistent for participants with asymmetrical hearing-loss. This result is coherent with previous studies that investigated the benefits of head-movements on front-back discrimination in young adults with normal-hearing (Perrett and Noble, 1997a,b; Iwaya et al., 2003) or with hearing-impairment (Mueller et al., 2014; Brimijoin and Akeroyd, 2016). Similarly, it corroborates the benefit of head-movements for sound localization in front space reported for young adults with normal-hearing (Brimijoin et al., 2013; Morikawa and Hirahara, 2013) or hearing-impairment (Coudert et al., 2022).

Our study extends to aging adults with ARHL the literature on the advantage of head-movements on sound localization. Although it has been documented that sound localization abilities decrease with advancing age (Dobreva and O'Neill, 2011), previous studies have mostly adopted a static-head approach when examining aging participants. To the best of our knowledge, the only exception to this wide-spread approach is represented by a study by Otte et al. (2013), in which they registered comparable localization performance in the horizontal dimension in young and older adults. Participants were exposed to an open-loop head-movement localization paradigm with sound sources varying horizontally and on the vertical plane. Target sounds were set to last 150 ms, precisely to ensure that the head-saccades toward the sound "always started after stimulus offset, which denied listeners potential acoustic feedback during their response" (Otte et al., 2013; p. 264). Yet, participants were free to move their head during the task and encoded sound position within a reference-frame that served head-movement. Although Otte et collaborators allowed head-movement in aging adults, the stimuli were too short to allow active listening experience (i.e., moving the head during the sound emission). Furthermore, the authors did not compare older adults' performance during head-fixed vs. free to move condition.

In the present study, we compared directly the two listening conditions. We did not manipulate directly the possibility of exploiting binaural and monaural cues by altering sounds frequency. However, during the active listening condition, participants were free to explore the acoustic space as they wanted, for a relatively long period of time (5 s). In the following paragraph we discuss the possible reasons subtending benefit of spontaneous head movement while listening.

The first set of explanation is related to the more peripheral consequences of moving the head. Wallach was the first to suggest that head rotations along the vertical axis can reduce the "cone of confusion" by 50% during front-back discrimination of sounds (Wallach, 1940). Wightman and Kistler (1999) proved that this benefit can be experienced even when sound sources change position while listeners maintain a static head-posture. This suggests that head-movement benefits may partly reflect the richer auditory cues available to the ears as the peripheral input becomes dynamic (see also Thurlow and Runge, 1967; Perrett and Noble, 1997a,b; Kato et al., 2003; McAnally and Martin, 2014). In this perspective, ARHL participants tested in this study struggled to discriminate front-back in head-fixed listening condition, likely as a consequence of their impoverished monaural spectral cues. Participants could have benefited from head-movements because turning the head changed front-back discrimination from a purely monaural task, to a task that could be solved exploiting binaural auditory cues. Another possibility is that moving the head introduced greater dynamicity in the monaural cues available at the ears. In support of this second, additional, interpretation we observed that asymmetrical ARHL participants improved in the 3D sound localization in front space task specifically in the vertical dimension. In other words, it appears that when localizing sounds head-movements allowed them to better exploit the monaural auditory cues needed for discriminating sound position in elevation.

In addition to these explanations based on changes occurring at the peripheral level, it is important to consider that head-movements are a paradigmatic example of active listening and can also reflect self-regulating strategies. In this respect, the interpretation of any head-movement related benefit becomes more cognitive, i.e., related to predictive behaviors that participants put in place when aiming to solve perceptual uncertainties. For asymmetrical ARHL participants, turning the head may have been an intentional strategy to exploit the head-shadow effect, maximizing sound intensity at the best ear (see also Valzolgher et al., 2020b, 2022b).

In conclusion, our findings extend to ARHL the literature on the advantage of head-movements on sound localization, and provide initial evidence that this benefit may be influenced by the disparity of auditory cues at the two ears. These results could be exploited when planning specific interventions in different hearing-impaired populations, and particularly point to the possibility of taking advantage of both acoustic benefit of head-movements and active behavioral strategies when promoting spatial hearing skills. Moreover, the present study underlines the importance of promoting more ecological scenarios, in order to consider active listening conditions during the evaluation of auditory abilities.

Given the limited number of participants enrolled in this research our findings provide only preliminary evidence. Future studies should examine the effect of active listening across different severity of hearing loss, in large scale studies. They may also compare participants with similar hearing-impairment but different age (young vs. aged) to examine the possible contributions of aging on the observed effects. Furthermore, it would be important to test the effect of moving the head as a function of sound features, and particularly examine these effects with more ecological sounds (e.g., speech).

Finally, although in the present study head-movements were implemented spontaneously, active head-orienting to sounds could be trained. Studies in this direction have already been conducted in normal hearing young adults with one ear plugged, to simulate a unilateral hearing loss condition (i.e., Valzolgher et al., 2020b, 2022b) and in bilateral cochlear implant users (Valzolgher et al., 2022a). A relevant future direction for research would be to test training paradigms to promote effective behavioral strategies during sound localization even in ARHL.

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: osf.io/57chk.

Ethics statement

The studies involving human participants were reviewed and approved by Research Ethics Committee, University of Trento (2019–037). 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

EGe, DS, GF, and FP contributed to conception and design of the study. EGe, GV, and CV organized the database. EGe, FP, and CV performed the statistical analysis. EGe and CV wrote the first draft of the manuscript. EGe, EGi, AF, FP, and CV wrote sections of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: FP was supported by a grant of the Agence Nationale de la Recherche (ANR-16-CE17-0016, VIRTUALHEARING3D, France), by a prize of the Foundation Medisite (France), by the Neurodis Foundation (France), and by a grant from the Italian Ministry for Research and University (MUR, PRIN 20177894ZH). CV was supported by a grant of the Università Italo-Francese/Université Franco-Italienne, the Zegna Founder's Scholarship and Associazione Amici di Claudio Demattè. EGe, FP, and CV were supported by a grant of the Velux Stiftung (n. 1439). FP, CV, and AF were supported by a grant of the Agence Nationale de la Recherche (ANR-16-CE17-0016, VIRTUAL-HEARING3D, France) and by a prize of the Fondation Medisite (France). The study was supported by the IHU CeSaMe ANR-10-IBHU-0003 and it was performed within the framework of the LABEX CORTEX (ANR-11- LABX-0042) of Université de Lyon.

Acknowledgments

The authors are grateful to all participants who took part in the study despite the difficulties related to the COVID-19 pandemic. We thank Giordana Torresani for drawing Figure 1, and two reviewers for their constructive comments on a previous version of the manuscript.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2022.1026056/full#supplementary-material

References

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). *Fitting Linear Mixed-Effects Models Using Lme4*. arXiv preprint:1406, 5823. doi: 10.18637/jss.v067.i01

Brimijoin, W. O., and Akeroyd, M. A. (2016). The effects of hearing impairment, age, and hearing aids on the use of self-motion for determining front/back location. *J. Am. Acad. Audiol.* 27, 588–600. doi: 10.3766/jaaa.15101

Brimijoin, W. O., Boyd, A. W., and Akeroyd, M. A. (2013). The contribution of head movement to the externalization and internalization of sounds. *PLoS ONE* 8, e83068. doi: 10.1371/journal.pone.0083068

Brimijoin, W. O., McShefferty, D., and Akeroyd, M. A. (2012). Undirected head movements of listeners with asymmetrical hearing impairment during a speech-in-noise task. *Hear. Res.* 283, 162–168. doi: 10.1016/j.heares.2011.10.009

Clark, J. G. (1981). Uses and abuses of hearing loss classification. *Asha* 23, 493–500.

Conti, S., Bonazzi, S., Laiacona, M., and Masina, M. (2015). Montreal cognitive assessment (MoCA)-Italian version: regression based norms and equivalent scores. *Neurol. Sci.* 36, 209–214. doi: 10.1007/s10072-014-1921-3

Coudert, A., Gaveau, V., Gatel, J., Verdelet, G., Salemme, R., Farnè, A., et al. (2022). Spatial hearing difficulties in reaching space in bilateral cochlear implant children improve with head movements. *Ear Hear.* 43, 192. doi: 10.1097/AUD.0000000000001090

Crespo, A. B., Idrovo, G. G., and Rodrigues, N. (2016). "A virtual reality UAV simulation with body area networks to promote the elders life quality," in *2016 1st International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW)* (IEEE), p. 1–7. doi: 10.1109/TISHW.2016.7847780

Davis, A. C. (2019). Hearing loss: rising prevalence and impact. *Bull. World Health Organ.* 97, 646. doi: 10.2471/BLT.19.224683

Dobreva, M. S., and O'Neill, W. E. (2011). Influence of aging on human sound localization. *J. Neurophysiol.* 105, 2471–2486. doi: 10.1152/jn.00951.2010

Eddins, A. C., Ozmeral, E. J., and Eddins, D. A. (2018). How aging impacts the encoding of binaural cues and the perception of auditory space. *Hear. Res.* 369, 79–89. doi: 10.1016/j.heares.2018.05.001

Fox, J., and Weisberg, S. (2020). Using car and effects functions in other functions. *Published Online* 1–5.

Gaveau, V., Coudert, A., Salemme, R., Koun, E., Desoche, C., Truy, E., et al. (2022). Benefits of active listening during 3D sound localization. *Exp. Brain Res.* 1–17. doi: 10.1007/s00221-022-06456-x

Hofman, P. M., and Van Riswick, J. G., and Van Opstal, A. J. (1998). Relearning sound localization with new ears. *Nat. Neurosci.* 1, 417–421. doi: 10.1038/1633

Iwaya, Y., Suzuki, Y., and Kimura, D. (2003). Effects of head movement on front-back error in sound localization. *Acoustical Sci. Technol.* 24, 322–324. doi: 10.1250/ast.24.322

Kato, M., Uematsu, H., and Kashino, M. (2003). The effect of head motion on the accuracy of sound localization. *Acoustical Sci. Technol.* 24, 315–317. doi: 10.1250/ast.24.315

Kumpik, D. P., and King, A. J. (2019). A review of the effects of unilateral hearing loss on spatial hearing. *Hear. Res.* 372, 17–28. doi: 10.1016/j.heares.2018.08.003

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. (2017). lmertest package: tests in linear mixed effects models. *J. Stat. Software* 82, 1–26. doi: 10.18637/jss.v082.i13

Macpherson, E. A., and Middlebrooks, J. C. (2002). Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited. *J. Acoust. Soc. Am.* 111, 2219–2236. doi: 10.1121/1.1471898

Majdak, P., Goupell, M. J., and Laback, B. (2010). 3-D localization of virtual sound sources: effects of visual environment, pointing method, and training. *Attention Percept. Psychophys.* 72, 454–469. doi: 10.3758/APP.72.2.454

McAnally, K. I., and Martin, R. L. (2014). Sound localization with head movement: implications for 3-d audio displays. *Front. Neurosci.* 8, 210. doi: 10.3389/fnins.2014.00210

Middlebrooks, J. C. (2015). "Sound localization," in *The Human Auditory System: Fundamental Organization and Clinical Disorders*, vol. 129. doi: 10.1016/B978-0-444-62630-1.00006-8

Morikawa, D., and Hirahara, T. (2013). Effect of head rotation on horizontal and median sound localization of band-limited noise. *Acoustical Sci. Technol.* 34, 56–8. doi: 10.1250/ast.34.56

Mueller, M. F., Meisenbacher, K., and Lai, W. K. (2014). Sound localization with bilateral cochlear implants in noise: how much do head movements contribute to localization? *Cochlear Implants Int.* 15, 36–42. doi: 10.1179/1754762813Y.0000000040

Noble, W., and Ter-Horst, K., and Byrne, D. (1995). Disabilities and handicaps associated with impaired auditory localization. *J. Am. Acad. Audiol.* 6, 129–140.

Otte, R. J., Agterberg, M. J., Van Wanrooij, M. M., and Snik, A. F. (2013). Age-related hearing loss and ear morphology affect vertical but not horizontal sound-localization performance. *J. Associat. Res. Otolaryngol.* 14, 261–273. doi: 10.1007/s10162-012-0367-7

Pastore, M. T., Natale, S., and Yost, W. (2018). Head movements allow listeners bilaterally implanted with cochlear implants to resolve front-back confusions. *Ear Hear.* 39, 1224. doi: 10.1097/AUD.0000000000000581

Perrett, S., and Noble, W. (1997a). The contribution of head motion cues to localization of low-pass noise. *Percept. Psychophys.* 59, 1018–1026. doi: 10.3758/BF03205517

Perrett, S., and Noble, W. (1997b). The effect of head rotations on vertical plane sound localization. *J. Acoust. Soc. Am.* 102, 2325–2332. doi: 10.1121/1.419642

Pollack, I., and Rose, M. (1967). Effect of head movement on the localization of sounds in the equatorial plane. *Percept. Psychophys.* 2, 591–596. doi: 10.3758/BF03210274

Rakerd, B., Vander Velde, T. J., and Hartmann, W. M. (1998). Sound localization in the median sagittal plane by listeners with presbycusis. *J. Am. Acad. Audiol.* 9, 466–479.

Savel, S., Drake, C., and Rabau, G. (2009). Human auditory localisation in a distorted environment: water. *Acta Acustica United Acustica* 95, 128–141. doi: 10.3813/AAA.918134

Thurlow, W. R., and P. S. Runge. (1967). Effect of induced head movements on localization of direction of sounds. *J. Acoust. Soc. Am.* 42:480–488. doi: 10.1121/1.1910604

Valzolgher, C., Alzhaler, M., Gessa, E., Todeschini, M., Nieto, P., Verdelet, G., et al. (2020a). The impact of a visual spatial frame on real sound-source localization in virtual reality. *Curr. Res. Behav. Sci.* 1, 100003. doi: 10.1016/j.crbeha.2020.100003

Valzolgher, C., Gatel, J., Bouzaid, S., Grenouillet, S., Todeschini, M., Verdelet, G., et al. (2022a). Reaching to sounds improves spatial hearing in bilateral cochlear implant users. *Ear Hear.* 10–1097. doi: 10.1097/AUD.0000000000001267

Valzolgher, C., Todeschini, M., Verdelet, G., Gatel, J., Salemme, R., Gaveau, V., et al. (2022b). Adapting to altered auditory cues: generalization from manual reaching to head pointing. *PLoS ONE* 17, e0263509. doi: 10.1371/journal.pone.0263509

Valzolgher, C., Verdelet, G., Salemme, R., Lombardi, L., Gaveau, V., Farn,é, A., et al. (2020b). Reaching to sounds in virtual reality: a multisensory-motor approach to promote adaptation to altered auditory cues. *Neuropsychologia* 149, 107665. doi: 10.1016/j.neuropsychologia.2020.107665

Wallach, H. (1940). The role of head movements and vestibular and visual cues in sound localization. *J. Exp. Psychol.* 27, 339. doi: 10.1037/h0054629

Wightman, F. L., and Kistler, D. J. (1999). Resolution of front–back ambiguity in spatial hearing by listener and source movement. *J. Acoust. Soc. Am.* 105, 2841–2853. doi: 10.1121/1.426899

frontiers | Frontiers in Human Neuroscience

TYPE Original Research
PUBLISHED 16 November 2022
DOI 10.3389/fnhum.2022.988644

OPEN ACCESS

EDITED BY
Elena Nava,
University of Milano-Bicocca, Italy

REVIEWED BY
Giovanna Mioni,
University of Padua, Italy
Guido Marco Cicchini,
National Research Council (CNR), Italy

*CORRESPONDENCE
Sara Incao
sara.incao@iit.it

SPECIALTY SECTION
This article was submitted to
Sensory Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 07 July 2022
ACCEPTED 24 October 2022
PUBLISHED 16 November 2022

CITATION
Incao S, Mazzola C and Sciutti A
(2022) The impact of early aging on
visual perception of space and time.
Front. Hum. Neurosci. 16:988644.
doi: 10.3389/fnhum.2022.988644

The impact of early aging on visual perception of space and time

Sara Incao[1,2]*, Carlo Mazzola[2,3] and Alessandra Sciutti[1]

[1]Cognitive Architecture for Collaborative Technologies Unit, Italian Institute of Technology (IIT), Genoa, Italy, [2]Dipartimento di Informatica, Bioingegneria, Robotica e Ingegneria dei Sistemi (DIBRIS), University of Genoa, Genoa, Italy, [3]Department of Robotics, Brain and Cognitive Science, Italian Institute of Technology (IIT), Genoa, Italy

Visual perception of space and time has been shown to rely on context dependency, an inferential process by which the average magnitude of a series of stimuli previously experienced acts as a prior during perception. This article aims to investigate the presence and evolution of this phenomenon in early aging. Two groups of participants belonging to two different age ranges (Young Adults: average age 28.8 years old; Older Adults: average age 62.8 years old) participated in the study performing a discrimination and a reproduction task, both in a spatial and temporal conditions. In particular, they were asked to evaluate lengths in the spatial domain and interval durations in the temporal one. Early aging resulted to be associated to a general decline of the perceptual acuity, which is particularly evident in the temporal condition. The context dependency phenomenon was preserved also during aging, maintaining similar levels as those exhibited by the younger group in both space and time perception. However, the older group showed a greater variability in context dependency among participants, perhaps due to different strategies used to face a higher uncertainty in the perceptual process.

KEYWORDS

context dependency, visual perception, temporal perception, spatial perception, early aging, regression to the mean, Bayesian models, central tendency

Introduction

The perception of space and time is very relevant for everyday life: consider the number of spatial and temporal estimations made when driving a car or when crossing a road. To make inferences about the world, humans base their predictions on past experience. Our knowledge of phenomena previously observed is the key to face the uncertainty derived by sensory experience. Hence, the perceptual process can be seen as an integration between the information coming from the senses and prior experience organized in internal models that act as priors. An example of such integration is represented by context dependency, also known as central tendency (Helmholtz, 1866; Hollingworth, 1910; Jazayeri and Shadlen, 2010; Karaminis et al., 2016; Roach et al., 2017). This phenomenon describes the way the predictive model is formed while perceiving a series of stimuli: the perception of each stimulus is influenced by the previous ones, so that the overall perception of the whole series of stimuli gravitates toward a mean magnitude. For instance, when we are shown several long segments and then we are asked to judge the length of another, shorter, one, we will perceive this one as longer than its actual size. This is because the perception is based on the prior, that can correspond to the average of the stimuli perceived before.

Context dependency has been observed in different perceptual domains and across different senses. These include auditory perception of time intervals (Jazayeri and Shadlen, 2010; Cicchini et al., 2012; Karaminis et al., 2016; Roach et al., 2017), visual perception of lengths (Sciutti et al., 2014; Mazzola et al., 2020), of points in space (Bejjanki et al., 2016), of categories (Huttenlocher et al., 2000) and of objects (Kersten and Yuille, 2003), and visual speed perception (Weiss et al., 2002; Stocker and Simoncelli, 2006). It has been shown that the integration process between sensory inputs and past experience can be modeled in Bayesian terms (Cicchini et al., 2012; Sciutti et al., 2014; Karaminis et al., 2016). According to this model, the more uncertain the sensory input is, the more relative weight is given to prior knowledge. Therefore, a great uncertainty in the sensory input should lead to a higher reliance on the prior, i.e., on stimulus history, and hence to a greater regression to the mean. Although some studies have investigated context dependency during development (Sciutti et al., 2014; Karaminis et al., 2016; Hallez et al., 2019), to our knowledge the phenomenon has been explored in aging in the temporal domain only by Gu et al. (2016). The present study aims at bridging this gap by assessing whether and how context dependency in the visual perception of space and time is influenced by the early phases of aging. Gu et al. (2016) focused their research on temporal memory as a function of temporal context dependency. To this aim, they designed the reproduction tasks with stimuli ranging from 7 s to 14 s, a time interval which is beyond the threshold of the psychological present (Fraisse, 1984). In our study, instead, we wanted to estimate how context dependency affected participants' perception by using shorter stimuli ranging from 1.27 s to 1.8 s as in Karaminis et al. (2016) with an additional discrimination task as a perceptual acuity assessment. In particular, we focused on changes occurring in people still active in society. For this reason, we selected participants with a mean age of about 60 years, who were still active in their work or in other activities that require a high perceptual and cognitive load. Regarding the visuo-spatial perception, and particularly the ability to visually discriminate lengths, differences between younger and older adults are reportedly not significant (Norman et al., 2014; Billino and Drewing, 2018). Also, Latham and Barrett (1998) found no age-related effect in a spatial discrimination task. In their experiment, participants had to decide whether the separation of the stimuli (distance between two white luminance patches) presented in the first interval was longer or smaller than the one presented in the second interval. However, as Faubert (2002) pointed out, visuo-spatial perceptual processes are affected by aging in case of increasing cognitive demand required by more complex tasks, such as delayed matching tasks or the processing of more than a single attribute per stimulus. This result is consistent with Lemay et al. (2004), which found that the difference between older and younger participants is visible only in the condition that required the stimulus to be remembered before executing the movement to reproduce the length of the target. They hypothesize that the information about target location was no longer available in the iconic memory of older participants having the stimulus been presented 1.5 s before the reproduction phase. The ability to visually discriminate distances seems to be preserved with age growing unless the tasks involve cognitive demands, in particular working memory, or specific mechanisms of integration, a circumstance that could be traced to the processing of multiple attributes of a single stimulus.

As for what concerns visuo-temporal perception, one of the most evident differences with increasing age is the higher variability in the elderly's answers (Wittmann and Lehnhoff, 2005; Turgeon et al., 2016; Lamotte and Droit-Volet, 2017). For instance, Turgeon et al. (2016) suggest that greater is the age, more variable are the tapping rates in different tapping tasks. In regard to time sensitivity, it has been demonstrated that perceptual acuity gradually declines with increasing age (Lamotte and Droit-Volet, 2017; Scurry et al., 2019; Mioni et al., 2021). Specifically, Lamotte and Droit-Volet (2017) found a decline in time sensitivity in the older group (76–81 years) with a bisection task. Mioni et al. (2021) assessed a worsening in perceptual acuity with a time discrimination task employing comparison intervals of 0.5 or 1.5 s. They found that when the standard stimulus was 1.5 s (similar to the design of our study: 1.535 s), the main differences already occurred from the age of 45. Gu et al. (2016) and Mioni et al. (2020) underlined a decline in time reproduction accuracy. In Mioni et al. (2020), besides an increased variability, older subjects showed a general tendency to underestimate their temporal judgments when asked to reproduce a time interval but to overestimate them in time production task. By contrast, Gu et al. (2016) found an effect of context dependency leading the aged group to a higher accuracy bias of the reproduced duration with respect to the stimuli. In addition, the role of cognitive functions is believed to account for the age-related changes in temporal perception. For example, mechanisms of attention and working memory that should decline with advanced age, play a fundamental role in evaluating the changes in perception of time for older adults (age range: 60–80 years old; Baudouin et al., 2006; Bartholomew et al., 2015; Brown et al., 2015).

The objective of this study was therefore to investigate how visual perception of space and time evolves in early aging in terms of perceptual acuity and use of priors. From previous literature and the Bayesian model of Context Dependency, the expectation is to find a stronger regression to the mean for the older adults in visual time perception to compensate an increased sensory uncertainty. Conversely, similar degrees of context dependency are expected for space perception between the two different age groups.

Materials and methods

The aim of the study was to evaluate whether the mechanism of context dependency in visual perception of space and time undergoes a change throughout life. To address this question, we asked participants from two different age groups to perform six tasks. Three tasks were designed to investigate the perception of space and three to assess the perception of time.

Participants

Forty-seven participants in total were recruited for this study. Twenty-five participants were classified as "*Young adults*" (YA), 12 males, 13 females (*M* = 28.8 years old, SD = 4.6). Twenty-two participants were classified as "*Old adults*" (OA) nine males, 13 females (*M* = 62.8 years old, SD = 4.1). The age range in the OA group was selected to include participants who could potentially exhibit age-related decline in visual perception, while maintaining good motor and cognitive abilities. In the *OA* group, one participant was excluded for the impossibility to complete the task, leaving a sample of 21 participants (eight male and 12 female). The study was approved by the regional ethical committee and all participants provided written informed consent before participating. All participants had normal or corrected-to-normal vision.

Task design

The experiment was divided in two conditions: space and time. Both conditions comprised three tasks. The order of conditions and tasks was randomized between participants. The experiment was performed in rooms lighted only with a lamp with a 11.5-Watt and 92 lm/W bulb placed near the screen, but pointing at the wall in front of the participant. The low-light condition was designed to avoid reflections on the screen. In both conditions, the participant sat on a chair with no wheels at a distance of 60 cm from the screen that was placed on a table (height 75 cm). The experiment was programmed and run with MATLAB 2019a and Psychtoolbox on a Windows 10 pc (Dell Inspiron 14 5000 2-1). In the space condition tasks, the stimuli were shown on a touchscreen ELO 2002L 20" monitor (resolution of 1,920 × 1,080 px for an active area of 436.9 mm × 240.7 mm, at a frequency of 60 Hz and Response Time of 0.02 s), whereas in the time condition tasks stimuli were shown directly on the Dell laptop screen (resolution of 1,920 × 1,080 px for an active area of 309.35 mm × 173.99, at a frequency of 60 Hz).

Space condition tasks

Pointing execution error—control task

Since the experiment included a reproduction task, we wanted to verify whether there were any significant differences between the two age groups in terms of motor abilities. For this reason, we designed a task where the stimuli were always visible and the participants were instructed to reach and touch them with their finger. For 50 trials, participants saw a red dot equal to the ones of the other spatial tasks, appearing in a random position on the screen. The number of possible positions was five in total: upper left corner, upper right corner, lower left corner, lower right corner (all of them 2 cm distant from the frame of the screen) and center of the screen. The participant had to touch the screen at the center of the red dot with as much precision as possible. The dot remained visible until the participant had completed the touch. The accuracy of the touch was measured as the distance between the center of the red dot and the touch of the participant.

Space discrimination task

This task was designed to assess the perceptual acuity of participants. Three red dots of 1 cm diameter were shown simultaneously for 0.4 s on a white straight line crossing the screen at its central height. When the stimulus disappeared, participants had to judge whether the longest segment was the first—delimited by the first dot and the second one—or the second—delimited by the second dot and the third—pressing respectively the "1" or "2" button on the keyboard (see **Figure 1A**). One of the lengths was always 10 cm (standard) while the other (comparison) was showed according to the QUEST adaptive procedure (Watson and Pelli, 1983): starting value: 12.0 ± 3.6 cm (SD).

Space reproduction task

The set-up of this task was the same as the previous one, but the red dots appearing were two, determining a segment of a certain length. The participant was asked to reproduce the same length between the first and the second dot, touching the screen in a third point so that the distance between the first and the second dot was the same as the distance between the second dot and the touch of the participant (see **Figure 1B**). The first red dot appeared at a distance from the left border of the screen, ranging from 0.2 to 1.7 cm randomly selected. Six sets of 11 different lengths randomly shown were presented to each participant for 66 trials. The lengths were ranging from 6 to 14 cm, increasing each 0.8 cm as in (Sciutti et al., 2014; Mazzola et al., 2020). No clues about the correctness of the answers were given to participants.

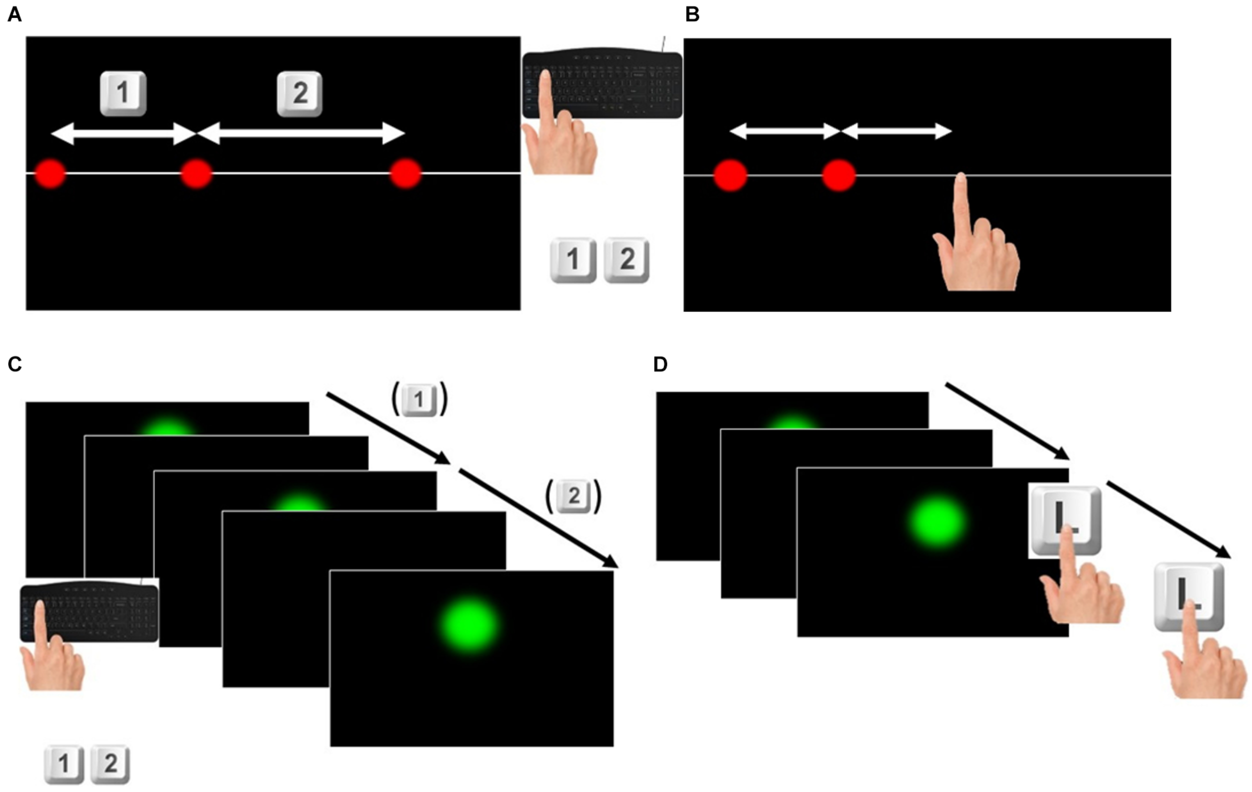


FIGURE 1
(A) Space discrimination task. For each trial participants were shown three red dots appearing simultaneously without any time interval between them. They were asked to judge which one was the longer and to press respectively the keys 1 or 2 on the keyboard if the longer was the one on the left or on the right. **(B)** Space reproduction task. For each trial participants were shown two red dots appearing consecutively without any time interval between them. They were asked to touch the touchscreen at the right of the second dot, to reproduce the distance between the two dots, by taking the second dot as reference. **(C)** Time discrimination task. For each trial participants were shown three green dots appearing at a certain time interval one from the other. They were asked to judge which one of the two time intervals was longer pressing respectively the keys 1 or 2 if the longer interval was the first (elapsed between the first and the second dot appeared) or the second (elapsed between the second and the third dot appeared). **(D)** Time reproduction task. For each trial participants were shown two green dots appearing at a certain time interval one from the other. They were asked to touch twice the letter L on the keyboard to reproduce the time interval between the two dots, pressing the start and the end of the time interval.

Time condition tasks

Rhythm synchronization task—control task

This task was performed in order to measure participants' ability to follow a constant rhythm, following a visual signal on the screen. The task consisted in following the rhythm marked by an intermittent green dot appearing on the screen for 50 trials. The green dot had a diameter of 2.2 cm and was placed 7.5 cm above the center of the screen. The participants were instructed to only look at the intermittent green dots for the first four appearances and internalize the rhythm. Then, they had to start pressing a keyboard key in order to synchronize the keypress with the appearance of the green dot.

Time discrimination task

This task was designed in a similar way to the space condition. Three green dots (2.2 cm diameter) appeared on the screen for 0.2 s defining two different time intervals: the first interval between the first and the second dot, the second interval between the second and the third dot (see **Figure 1C**). One interval—the standard—was constant (1.535 s) while the comparison interval was defined according to a QUEST adaptive procedure (Watson and Pelli, 1983): starting value: 1.7 ± 0.52 s (SD). The first dot was presented on the screen at a randomly varying time interval from the start of the trial, ranging from 1 s to 1.8 s. The number of trials performed by participants was not less than 50 but could vary according to the QUEST adaptive procedure.

Time reproduction task

The set-up of this task was similar to the previous one, but the dots (2.2 cm diameter, appearing on the screen for 0.2 s) were only two, therefore showing a single time interval. The participant had to reproduce this time interval pressing

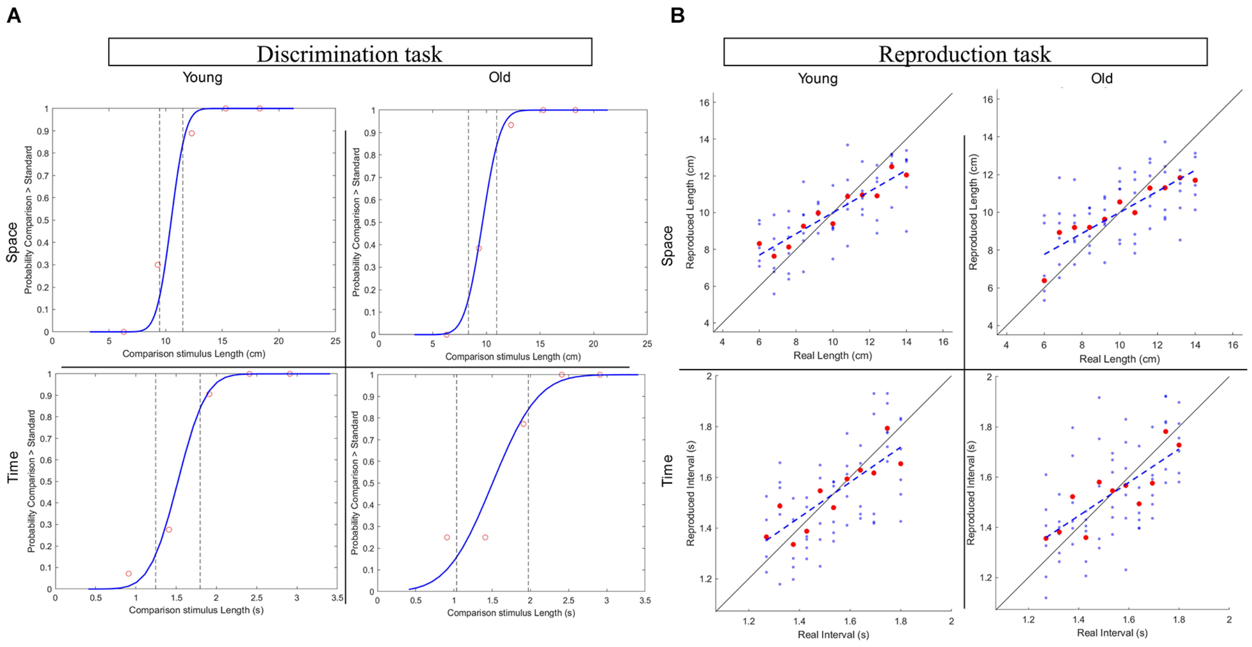


FIGURE 2
(A) Representative subjects' plots for the Discrimination task in the spatial and temporal conditions from the two age groups. **(B)** Representative subjects' plots for the Reproduction task in the spatial and temporal conditions from the two age groups.

twice the letter L on the keyboard, so as to simulate the start and end (see **Figure 1D**). In a similar way to the space reproduction task, a set of 11 different time intervals, ranging from 1.270 s to 1.8 s as in (Karaminis et al., 2016) was shown six times and randomized within each set of stimuli, for a total of 66 trials. The first dot appeared on the screen after a varying delay from the start of the trial ranging from 1 to 2 s.

Data analysis

Control tasks

For the pointing execution error task, we computed the distance between the stimulus shown on the screen (s) and the point touched by participants (r).

$$\text{Pointing err.} = \sqrt{(x_r - x_s)^2 + (y_r - y_s)^2}$$

For the rhythm synchronization task, we measured participants' ability to reproduce a rhythm by calculating the standard deviation of the time intervals indicated by the keypress.

$$\text{Rhythm Variability} = \sqrt{\frac{\sum_{i=1}^{N} (x_i' - \bar{x})^2}{N}}$$

where x_i' is the time interval between two consecutive keypresses. Since this measure was only relative to the temporal condition, a Wilcoxon Mann-Whitney test was used to compare the results of the two populations (YA, OA).

Discrimination tasks

For space and time condition, the differential threshold of each participant, that is the minimal difference between two lengths or time intervals that participants could reliably discriminate, was calculated as the standard deviation of the psychometric function (cumulative gaussian) fitted on the data of the discrimination tasks (see **Figure 2A**). Then, perceptual acuity was expressed as Weber Fraction (WF) measured as the ratio between the threshold and the standard stimulus (Cicchini et al., 2012; Karaminis et al., 2016).

Reproduction tasks

In the reproduction tasks, we evaluated both the average and the absolute perceptual bias of participants. Specifically, the offset was calculated by subtracting the average stimulus of all the trials ($\bar{S}$) from the average response ($\bar{R}$) to indicate participants tendency to overestimate or underestimate stimuli.

$$\text{Offset} = \bar{R} - \bar{S}$$

Well-established approaches were followed to estimate the degree of central tendency in spatial and temporal perception (Cicchini et al., 2012; Sciutti et al., 2014). As a direct measure of central tendency, we computed the regression index for each of the reproduction tasks (*Space Reproduction, Time Reproduction*) as the difference in slope between the identity line (ideal correct responses on stimuli) and the best linear fit of the responses given by the participant (participant's responses plotted against the correspondent stimuli), see **Figure 2B**. Moreover, we also calculated the overall perceptual error (RMSE) as the root mean square between the accuracy error (BiasCD) and the precision error (CV). For each: (i) of the 11 stimuli of the Reproduction tasks, the degree of accuracy (BiasCD) results from the difference between the responses average (R_{Mi}) and the corresponding stimulus (S_i), normalized for the average stimulus ($\bar{S}$). The precision error is computed as the coefficient of variation (CV), namely the ratio between the standard deviation of the responses of a stimulus and the average stimulus.

$$BiasCD_i = \frac{|R_{Mi} - S_i|}{\bar{S}}$$

$$\mathrm{CV}_i = \frac{\sqrt{\frac{\Sigma\left(R'_i - \bar{R'_i}\right)^2}{N}}}{\bar{S}}$$

$$\mathrm{RMSE}_i = \sqrt{BiasCD_i^2 + \mathrm{CV}_i^2}$$

Following previous studies on context dependency (Cicchini et al., 2012; Sciutti et al., 2014; Karaminis et al., 2016; Mazzola et al., 2022), all these errors are calculated for each participant after subtracting from the participants' responses (R) their Offset. In this way, the Offset is considered a perceptual offset caused by the individual tendency to perceive stimuli as greater or lower, independently from the stimulus history.

$$\mathrm{R}' = \mathrm{R} - \mathrm{Offset}$$

Seven Linear Mixed Effect Models have been used to compare the effects of the conditions (Space and Time) and of the population (YA and OA) on the values of the Weber Fraction, the Offset, the Regression Index, and the three perceptual errors connected to context dependency (BiasCD, CV, and RMSE). These statistical analyses were conducted using R software (R i386 4.0.3) and specific libraries for Linear Mixed Effect Models (Kuznetsova et al., 2017; Lüdecke, 2021). The above mentioned parameters (Weber Fraction, Offset, Regression Index, Bias CD, CV, RMSE) were inserted as dependent variable, the condition (space and time), the age (*YA, OA*), and their interaction as predictors, and the subjects as random effect.

Results

In this experiment we wanted to observe whether the visual perception of space and time and the central tendency mechanism supporting it are influenced by early aging in two representative age groups: YA and OA. In particular, we assessed potential variations in perceptual acuity and whether they had an impact on the use of prior knowledge in perception and on the participants' perceptual bias.

Control tasks

As regards the spatial control task, no significant difference was found in the pointing error when comparing the YA group (*Mdn* = 0.195 cm) with the OA group (*Mdn* = 0.243 cm) in a Mann-Whitney test: $U_{(25,21)} = 262.50$, $z = -1.61$, $p = 0.107$. This means that the basic motor abilities required to execute the spatial tasks were not significantly different with increasing age.

The rhythm synchronization task provided a measure of participants' ability to synchronize to a given rhythm. To investigate this, it was measured the variability of time intervals reproduced by participant while attempting to follow the predefined rhythm. A high variability indicates participant's difficulty in keeping up the pace with the rhythmic stimulus that was shown. Due to a technical problem, the data of two participants of the YA group for the rhythm synchronization task were not saved. A Mann-Whitney test on the remaining sample did not reveal any difference between the YA group (*Mdn* = 0.064 s) and the OA group (*Mdn* = 0.082 s): $U_{(23,21)} = 295$, $z = -1.245$, $p = 0.213$.

Perceptual acuity

The results of the discrimination tasks revealed that the Weber Fraction of the two populations of our study differs significantly only in the time condition (see **Figure 3** for data visualization, **Table 1** for means and **Table 2** for statistics). Indeed, the Linear Mixed Effect Model on the WF showed that the temporal perception threshold is significantly higher in the OA group (YA-OA: $B = -0.071$, $t = -2.998$, $p = 0.004$). The same test does not reach significance for the spatial condition (YA-OA: $B = -0.04$, $t = -1.673$, $p = 0.098$), even though the trend is the same (see **Table 1**). In addition, from the same Linear Mixed Effect Model, in both YA and OA groups, the comparison between the Weber Fractions of space and time condition (Space-Time) resulted significantly different (YA: $B = -0.084$, $t = -4.649$, $p < 0.001$; OA: $B = -0.116$, $t = -5.84$, $p < 0.001$). This shows that, independently of their age, participants found the temporal discrimination task more difficult than the spatial one. No effect of interaction between condition and age was

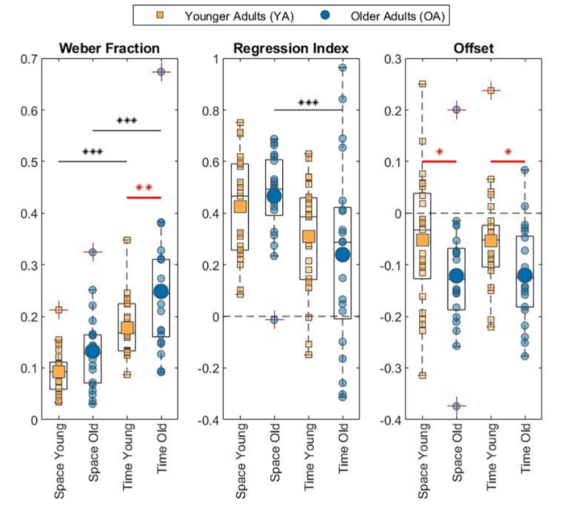


FIGURE 3
Box plot with individual data of the measures related to perceptual acuity (Weber Fraction), regression to the mean (Regression Index), and Offset. The asterisks mark statistical significance. Black ones denote difference between conditions. Red ones denote difference between age groups. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

found. Moreover, it was not found any correlation between conditions for either groups.

Context dependency

For both conditions, the two age groups exhibited a regression to the mean during the reproduction task (see **Table 1** for Mean and SD). Indeed, all Regression Indexes resulted significantly different from 0 in one-sample t-tests: YA Space: $t_{(24)} = 10.95$, Cohen's $d = 2.189$, $p < 0.001$; OA Space: $t_{(20)} = 12.57$, Cohen's $d = 2.742$, $p < 0.001$; YA Time: $t_{(24)} = 7.26$, Cohen's $d = 1.452$, $p < 0.001$; OA Time: $t_{(20)} = 3.02$, Cohen's $d = 0.66$, $p = 0.007$.

Focusing on the difference between age groups, no significant RI variations were found (see **Figure 3** for data visualization, **Table 1** for means and **Table 2** for statistics) neither in the spatial perception, nor in the temporal one. Nevertheless, in the OA group, only for the temporal domain, RI increased significantly with growing age: $F_{(1,19)} = 5.27$, $R^2 = 0.22$, $p = 0.033$.

As regards the difference between conditions, the Linear Mixed Effect Model of Regression Index showed that in the OA population the regression index was significantly lower in the visual perception of time than space (Space-Time: $B = 0.227$, $t = 3.023$, $p = 0.003$). Even though in the YA group no difference was found across conditions, the trend was the same (Space-Time: $B = 0.116$, $t = 1.688$, $p < 0.095$).

Concerning the three measures of perceptual errors connected to context dependency (BiasCD, CV, RMSE), see

TABLE 1 Means and Standard Deviations of the perceptual measures explored in the study for the two age groups.

Age	**Weber fraction**		**Offset**		**Regression index**		**Bias context dependency**		**Coefficient of Variation (CV)**		**Root Mean Squared Error (RMSE)**	
	Space	Time	Space	Time	Space	Time	Space	Time	Space	Time	Space	Time
YA	$M = 0.093$	$M = 0.177$	$M = -0.053$	$M = -0.054$	$M = 0.426$	$M = 0.320$	$M = 0.119$	$M = 0.067$	$M = 0.114$	$M = 0.141$	$M = 0.165$	$M = 0.157$
	SD = 0.043	SD = 0.056	SD = 0.128	SD = 0.092	SD = 0.195	SD = 0.193	SD = 0.045	SD = 0.014	SD = 0.022	SD = 0.040	SD = 0.029	SD = 0.039
OA	$M = 0.133$	$M = 0.248$	$M = 0.121$	$M = -0.120$	$M = 0.467$	$M = 0.294$	$M = 0.131$	$M = 0.073$	$M = 0.121$	$M = 0.137$	$M = 0.174$	$M = 0.156$
	SD = 0.074	SD = 0.129	SD = 0.114	SD = 0.092	SD = 0.168	SD = 0.294	SD = 0.036	SD = 0.024	SD = 0.037	SD = 0.05	SD = 0.043	SD = 0.050

TABLE 2 Results of the linear Mixed Effect Models for the seven perceptual measures of the discrimination and the reproduction tasks.

Results of the 7 Linear Mixed Effect Models

		condition: space-*time*/age: YA-*OA*					**condition: time-*space*/age: OA-*YA***				
		Estimate	**St. Err.**	**df**	**T**	***p***	**Estimate**	**St. Err.**	**df**	**T**	***p***
Weber Fraction (WF)	Intercept	0.248	0.017	78,109	14.245	<0.001	0.093	0.016	78,109	5.828	<0.001
	Condition	−0.116	0.020	44,000	−5.844	**<0.001**	0.084	0.018	44,000	4.649	**<0.001**
	Age	−0.071	0.024	78,109	−2.998	**0.004**	0.040	0.024	78,109	1.673	0.098
	Age*Condition	0.031	0.027	44,000	1.167	0.249	0.031	0.027	44,000	1.167	0.249
Offset	Intercept	−0.120	0.024	88,000	−5.109	<0.001	−0.053	0.022	88,000	−2.443	0.017
	Condition	−0.001	0.033	88,000	−0.026	0.979	−0.001	0.031	88,000	−0.029	0.977
	Age	0.067	0.032	88,000	2.088	**0.040**	−0.069	0.032	88,000	−2.143	**0.035**
	Age*Condition	−0.002	0.045	88,000	0.039	0.969	0.002	0.045	88,000	0.039	0.969
Regression Index (RI)	Intercept	0.240	0.053	88,000	4.509	<0.001	0.426	0.049	88,000	8.751	0.000
	Condition	0.227	0.075	88,000	3.023	**0.003**	−0.116	0.069	88,000	−1.688	0.095
	Age	0.070	0.072	88,000	0.976	0.332	0.041	0.072	88,000	0.563	0.575
	Age*Condition	−0.111	0.102	88,000	−1.088	0.280	−0.111	0.102	88,000	−1.088	0.280
Bias Context Dependency (BiasCD)	Intercept	0.073	0.007	87,746	10.421	<0.001	0.119	0.006	87,746	18.579	<0.001
	Condition	0.058	0.010	44,000	6.074	**<0.001**	−0.052	0.009	44,000	−5.905	**<0.001**
	Age	−0.006	0.009	87,746	−0.618	0.538	0.012	0.009	87,746	1.289	0.201
	Age*Condition	−0.006	0.013	44,000	−0.488	0.628	−0.006	0.013	44,000	−0.488	0.628
Coefficient of Variation (CV)	Intercept	0.137	0.008	72,285	16.731	<0.001	0.114	0.008	72,285	15.192	<0.001
	Condition	−0.017	0.008	44,000	−1.977	0.054	0.027	0.008	44,000	3.422	**<0.001**
	Age	0.004	0.011	72,285	0.320	0.750	0.006	0.011	72,285	0.564	0.575
	Age*Condition	−0.010	0.012	44,000	−0.855	0.397	−0.010	0.012	44,000	−0.855	0.397
Root Mean Squared Error (RMSE)	Intercept	0.156	0.009	77,621	17.661	<0.001	0.165	0.008	77,621	20.377	<0.001
	Condition	0.018	0.010	44,000	1.805	0.078	−0.008	0.009	44,000	−0.929	0.358
	Age	0.000	0.012	77,621	0.041	0.967	0.009	0.012	77,621	0.751	0.455
	Age*Condition	−0.010	0.014	44,000	−0.703	0.486	−0.010	0.014	44,000	−0.703	0.486

The left column gives the statistics if taking as reference the temporal condition and the OA group, whereas the right column if taking as reference the spatial condition and the YA group. Significant *p* values for age, condition and age*condition predictors are in bold.

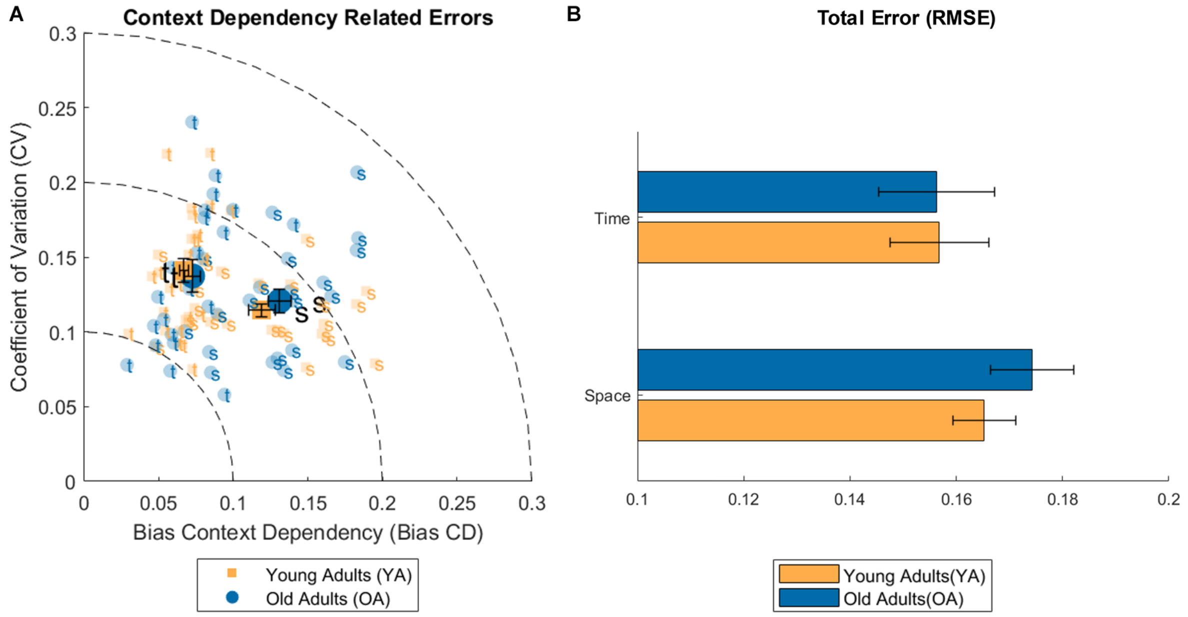


FIGURE 4
(A) Plot of the individual data for Bias and coefficient of variation (CV) related to the context dependency phenomenon (error bars represent the Standard Error of the mean). **(B)** Bar plot of the root mean squared error (RMSE) for the spatial and temporal condition and both age groups (error bars represent the Standard Error of the mean).

Figure 4, a significant difference between conditions has been found for the BiasCD (see Table 1 for means and Table 2 for statistics; Space-Time, YA: $B = 0.052$, $t = 5.905$, $p < 0.001$; OA: $B = 0.058$, $t = 6.074$, $p < 0.001$), revealing that both the age groups were more accurate in the temporal dimension. A significant variation, but in the opposite direction has been found for the OA group, also in the CV, demonstrating a loss in precision in the temporal domain (Space-Time, YA: $B = -0.027$, $t = -3.422$, $p = 0.001$). The same trend is visible also for the YA group (OA: $B = -0.017$, $t = -1.977$, $p = 0.054$). No difference between conditions has been found for the RMSE. Only the OA group exhibited a decreasing trend for the error in the temporal perception (Space-Time, OA: $B = 0.018$, $t = 1.805$ $p = 0.078$), although demonstrating a lower perceptual acuity (higher WF) in the discrimination task. No significant difference has been found between age groups for the three perceptual errors: neither in the spatial dimension, nor in the temporal one, despite the variation of perceptual acuity in time between younger and older adults. Moreover, no effect has been found for the interaction between age and condition, neither for the regression index, nor for the three perceptual errors related to context dependency. No correlation was found between conditions for either groups.

Perceptual reproduction offset

Independently from the phenomenon of context dependency, also the measures of the average Offset was calculated from the results of the reproduction tasks. All the Offset means resulted significantly different from 0 in both conditions and for both age groups in one-tailed *t*-tests and revealed that, on average, participants underestimated both the temporal and the spatial stimuli. Specifically, the Offset in space in the YA group: $t_{(24)} = -2.06$, $p = 0.05$, Cohen's $D = -0.41$, the Offset in time in the YA group: $t_{(24)} = -2.91$, $p = 0.008$, Cohen's $D = -0.58$, the Offset in space in the OA group: $t_{(20)} = -4.86$, $p < 0.001$, Cohen's $D = -1.06$, and the Offset in time in the OA group: $t_{(20)} = -5.99$, $p < 0.001$, Cohen's $D = -1.31$ (means and the SDs as reported in Table 1).

Linear Mixed Effect models assessed the variation of Offset for condition and age group. For the Offset, the statistical analysis showed a significant effect of Age in both conditions, which revealed the OA group perceived stimuli as shorter with respect to the YA group (see Figure 3 for data visualization, Table 1 for means and Table 2 for statistics; YA-OA, Space: $B = 0.069$, $t = 2.143$, $p = 0.035$; Time: $B = 0.067$, $t = 2.088$, $p = 0.040$). No effect has been found between conditions, nor for the interaction between Age and Condition. Moreover, it was not found any correlation between conditions for either groups.

Discussion

According to the aim of our research, we explored the effect of the early phases of aging on visual perception of space and time. The results indicate a general decline in perceptual acuity in the OA. Going deeper into the analysis of the two conditions, while for the spatial domain it is noticeable only with a slight decreasing trend, for the temporal domain, a significant difference between the two age groups emerged, which is also consistent with Mioni et al. (2021). In the selected temporal and spatial ranges, temporal discrimination resulted to be more difficult than spatial discrimination for both YA and OA groups, with Weber Fraction almost two times greater in the temporal than in the spatial condition. All participants had either normal or corrected-to-normal vision through the use of glasses. Such correction granted that all participants had the same potential capability to see the spatial stimuli presented in the task.

In this study, we also explored the phenomenon of context dependency in both domains of space and time. Such phenomenon intervenes when the information coming from the senses is uncertain and unprecise. Given the uncertainty of information, relying on our prior experience helps to reduce the variability of what is perceived. In accordance with previous research on the topic (Cicchini et al., 2012; Sciutti et al., 2014; Karaminis et al., 2016; Mazzola et al., 2022), we observed the mechanism of regression to the mean in both conditions, revealing a context dependency effect. In addition to the previous findings, this study found the phenomenon to be present also in an older population.

The close relation between visual acuity and use of priors has been found and modeled in a Bayesian fashion by previous studies (Cicchini et al., 2012; Sciutti et al., 2014; Karaminis et al., 2016; Mazzola et al., 2022). According to this model the decline of visual acuity we observed in the temporal domain with increasing age, would predict an increase in prior reliance. By contrast, the results of the time reproduction task did not reveal a higher regression index for the OA group compared to the younger group.

The transitional phase of the aging of the older population in this study may account for this result. Though the comparison between the two groups did not follow our predictions, age was shown to be a predictor for the RI within the OA group, such that in the temporal domain the older the subject, the higher the RI. From this perspective, one may speculate that if the decay of perceptual acuity is already visible from the age of 60 [in Mioni et al. (2021) already from 45], this same age is not sufficient to determine a consistent increase in the RI.

A different explanation is related to the mode of stimuli presentation. As Droit-Volet et al. (2008) demonstrated, differences in sensitivity of time, number, and length are only due to the sequential or nonsequential mode of presentation. In their experiment, when number and length were presented sequentially, as time is for its own nature (i.e., extended, with a duration), differences were leveled. Authors linked these findings with the higher attentional and cognitive resources required in the sequential presentation. Hence, these findings, combined with suggestions from previous literature (Faubert, 2002; Bartholomew et al., 2015; Lamotte and Droit-Volet, 2017), which indicate a worsening of elderly people's perceptual performances when a higher attentional and cognitive control is required, are consistent with our results from the discrimination task. In the spatial (nonsequential) condition, we did not find any significant difference between the two age groups, which instead was found in the temporal (sequential) condition.

Following Droit-Volet et al. (2008), since worse attentional and cognitive control causes a decline in perceptual acuity, we can hypothesize this receptive difficulty be present also in the reproduction task of our study, somehow affecting the prior formation. Here, the sequential mode of presentation may have influenced the phenomenon of context dependency at the level of stimuli reception. Considering the design of the reproduction tasks, in the spatial domain there is no temporal interval between the presentation of the first and second dot forming the stimulus. By contrast, in the time reproduction task, the onset and offset of the stimulus are spaced out by a certain temporal interval and therefore higher attentional and cognitive load is required. As a result, a great variability is visible among older participants in the time reproduction task and the phenomenon of context dependency, as indicated by the RI, is weaker than what expected. The case is different for the spatial condition. Here, when the stimuli presentation is nonsequential, context dependency seems not to be affected by growing age. In general, context dependency was shown to be present with age growing but its mechanism may be impacted in case of a higher attentional and cognitive demand.

The Offset is another measure to analyze the effect of aging in the reproduction of time. It represents the mean of perceptual bias, providing also the information about its direction. Both in the spatial and the temporal conditions, the data showed a general tendency to underestimate the stimuli amplitude with a negative Offset that becomes broader with increasing age. Regarding the spatial domain, it was not possible to connect the underestimation strategy found in the OA group with other spatial perceptual measures of this study. By contrast, the decrease of the Offset (underestimation) in the OA group is consistent with their decline in perceptual acuity. A feasible explanation for the temporal underestimation in the reproduction task might be offered by the hypothesis of the internal clock model (Grondin, 2010). This theory considers the presence of a main mechanism responsible for temporal estimation and explains the representation of time in terms of pulses emitted by an internal clock. Previous literature supports the idea of a slower internal clock in the elderly (Turgeon et al., 2016; Lamotte and Droit-Volet, 2017). A slower clock would be due to fewer pulses emitted and therefore counted, a phenomenon that in reproduction tasks results in an underestimation of durations (Perbal et al., 2002). In the context of this hypothesis, the rhythm synchronization task may provide an interesting insight. During this task, the visual feedback of the stimulus was always present on the screen providing a reference. Conversely, during the reproduction task participants could only rely on their internal clock to reproduce the time interval. Interestingly, no difference between the two age groups was found in the rhythm synchronization task. Hence, the visual perception of time seems to be affected by the increasing age only when no visual feedback is provided, i.e., whether participants can only rely on their internal clock. As explained by Marinho et al. (2018), the variation of the internal clock is strictly connected with the dopaminergic system. Consequently, the stronger underestimation of OA group in the temporal reproduction task may be motivated by the decline in dopaminergic modulation that is showed to be present in older age (Li et al., 2010).

The motivation at the basis of this study was to understand whether and how visual perception of space and time changes with the increase of age. We wanted to focus on the age range in which sensory perception already undergoes a significant degradation, but the life of a person is still very active and similarly demanding, in terms of spatio-temporal abilities, as a younger age. Around 60 years of age, indeed, most people are still working or performing a rich range of activities. For the spatial domain, our data show that the overall performance of the older group was quite similar to the younger adults in the context of visual perception of space, both in terms of spatial acuity and regression to the mean. Only a general tendency to underestimate spatial amplitudes in the reproduction task differentiated significantly young and older adults. Conversely, temporal visual acuity resulted significantly reduced in the older adult group, together with a similar general underestimation of temporal intervals in the reproduction task. These findings indicate that already early during aging, visual perception of time undergoes significant changes. Focusing on the phenomenon of context dependency, in general, it appears clear that, with increasing age, in the temporal domain, the phenomenon of context dependency occurs differently than what would be expected from a direct application of the Bayesian modeling. In particular, in face of a significant reduction of their perceptual acuity, participants in the early aging group did not increase their tendency to rely on their prior.

Although our study is not definitive with respect to the causes underlying such variation, two possible directions emerge. First, it might be the case that with an older population, the expectations of the Bayesian model will be confirmed. Hence, a study across three different ages, adding an older population might be of help. Second, the sequential mode of presentation, the role of attentional/cognitive effort and the relation between space and time need further investigation. Modifying the mode of presentation might reveal whether a sequential spatial task and a temporal task present the same perceptual difficulties. Regarding the discrimination, this could result in a difference between age groups in the perceptual acuity. Whereas in regard to the reproduction, it may lead to a deviation from the Bayesian predictions of context dependency. Furthermore, adding sequentiality in a spatial task could also help in determining whether both age groups are affected by similar cognitive challenges at the level of context dependency. In this perspective, a test for cognitive performance, which was not present in our experiment, could shed light on the impact of cognitive and attentive control regardless of participants' age. Leveling the cognitive difficulties among conditions might therefore be a possibility to deepen mechanisms connecting reception of stimuli, cognitive demand and the phenomenon of context dependency. Eventually, further research in this direction would be also crucial to understand whether in the elderly the deviation from Bayesian predictions is due to the higher cognitive and attentional demand, to different processes underlying spatial and temporal perception, or to other factors connected with aging.

Data availability statement

The 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 Regione Liguria. The participants provided their written informed consent to participate in this study.

Author contributions

SI and AS conceived and planned the experiments. SI and CM carried out the experiments and analyzed the data. All the authors contributed to the interpretation of the results. SI wrote the manuscript in consultation with AS and CM. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by a Starting Grant from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, G.A. No. 804388, wHiSPER.

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

References

Bartholomew, A. J., Meck, W. H., and Cirulli, E. T. (2015). Analysis of genetic and non-genetic factors influencing timing and time perception. *PLoS One* 10:e0143873. doi: 10.1371/journal.pone.0143873

Baudouin, A., Vanneste, S., Pouthas, V., and Isingrini, M. (2006). Age-related changes in duration reproduction: involvement of working memory processes. *Brain Cogn.* 62, 17–23. doi: 10.1016/j.bandc.2006.03.003

Bejjanki, V. R., Knill, D. C., and Aslin, R. N. (2016). Learning and inference using complex generative models in a spatial localization task. *J. Vis.* 16:9. doi: 10.1167/16.5.9

Billino, J., and Drewing, K. (2018). Age effects on visuo-haptic length discrimination: evidence for optimal integration of senses in senior adults. *Multisens. Res.* 31, 273–300. doi: 10.1163/22134808-00002601

Brown, S. W., Johnson, T. M., Sohl, M. E., and Dumas, M. K. (2015). Executive attentional resources in timing: effects of inhibitory control and cognitive aging. *J. Exp. Psychol. Hum. Percept. Perform.* 41, 1063–1083. doi: 10.1037/xhp0000078

Cicchini, G. M., Arrighi, R., Cecchetti, L., Giusti, M., and Burr, D. C. (2012). Optimal encoding of interval timing in expert percussionists. *J. Neurosci.* 32, 1056–1060. doi: 10.1523/JNEUROSCI.3411-11.2012

Droit-Volet, S., Clement, A., and Fayol, M. (2008). Time, number and length: similarities and differences in discrimination in adults and children. *Q. J. Exp. Psychol.* 61, 1827–1846. doi: 10.1080/17470210701743643

Faubert, J. (2002). Visual perception and aging. *Can. J. Exp. Psychol.* 56, 164–176. doi: 10.1037/h0087394

Fraisse, P. (1984). Perception and estimation of time. *Ann. Rev. Psychol.* 35, 1–37. doi: 10.1146/annurev.ps.35.020184.000245

Grondin, S. (2010). Timing and time perception: a review of recent behavioral and neuroscience findings and theoretical directions. *Atten. Percept. Psychophys.* 72, 561–582. doi: 10.3758/APP.72.3.561

Gu, B. M., Jurkowski, A. J., Shi, Z., and Meck, W. H. (2016). Bayesian optimization of interval timing and biases in temporal memory as a function of temporal context, feedback and dopamine levels in young, aged and Parkinson's disease patients. *Timing Time Percept.* 4, 315–342. doi: 10.1163/22134468-00002072

Hallez, Q., Damsma, A., Rhodes, D., van Rijn, H., and Droit-Volet, S. (2019). The dynamic effect of context on interval timing in children and adults. *Acta Psychol. (Amst)* 192, 87–93. doi: 10.1016/j.actpsy.2018.10.004

Helmholtz, H. (1866). *Handbuch der Physiologischen Optik*. Hamburg, Leipzig: Leopold Voss.

Hollingworth, H. L. (1910). The central tendency of judgment. *J. Philos. Psychol. Sci. Methods* 7, 461–469. doi: 10.2307/2012819

Huttenlocher, J., Hedges, L. v., and Vevea, J. L. (2000). Why do categories affect stimulus judgment. *J. Exp. Psychol. Gen.* 129, 220–241. doi: 10.1037//0096-3445.129.2.220

Jazayeri, M., and Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nat. Neurosci.* 13, 1020–1026. doi: 10.1038/nn.2590

Karaminis, T., Cicchini, G. M., Neil, L., Cappagli, G., Aagten-Murphy, D., Burr, D., et al. (2016). Central tendency effects in time interval reproduction in autism. *Sci. Rep.* 6:28570. doi: 10.1038/srep28570

Kersten, D., and Yuille, A. (2003). Bayesian models of object perception. *Curr. Opin. Neurobiol.* 13, 150–158. doi: 10.1016/s0959-4388(03)00042-4

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). ImerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13

Lamotte, M., and Droit-Volet, S. (2017). Aging and time perception for short and long durations: a question of attention. *Timing Time Percept.* 5, 149–167. doi: 10.1163/22134468-00002086

Latham, K., and Barrett, B. T. (1998). No effect of age on spatial interval discrimination as a function of eccentricity or separation. *Curr. Eye Res.* 17, 1010–1017. doi: 10.1021/acs.jcim.2c00847

Lemay, M., Gagnon, S., and Proteau, L. (2004). Manual pointing to remembered targets. but also in a remembered visual context. *Acta Psychol. (Amst)* 117, 139–153. doi: 10.1016/j.actpsy.2004.06.002

Li, S. C., Lindenberger, U., and Bäckman, L. (2010). Dopaminergic modulation of cognition across the life span. *Neurosci. Biobehav. Rev.* 34, 625–630. doi: 10.1016/j.neubiorev.2010.02.003

Lüdecke, D. (2021). *sjstats: Statistical Functions for Regression Models (Version 0.18.1)*. Available online at: https://CRAN.R-project.org/package=sjstats.

Marinho, V., Oliveira, T., Rocha, K., Ribeiro, J., Magalhães, F., Bento, T., et al. (2018). The dopaminergic system dynamic in the time perception: a review of the evidence. *Int. J. Neurosci.* 128, 262–282. doi: 10.1080/00207454.2017.1385614

Mazzola, C., Aroyo, A. M., Rea, F., and Sciutti, A. (2020). "Interacting with a social robot affects visual perception of space," in *HRI'20: Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY, USA: Association for Computing Machinery), 549–557.

Mazzola, C., Rea, F., and Sciutti, A. (2022). Shared perception is different from individual perception a new look on context dependency. *IEEE Trans. Cogn. Dev. Syst.* doi: 10.1109/TCDS.2022.3185100

Mioni, G., Cardullo, S., Ciavarelli, A., and Stablum, F. (2021). Age-related changes in time discrimination: the involvement of inhibition, working memory and speed of processing. *Curr. Psychol.* 40, 2462–2471. doi: 10.1007/s12144-019-00170-8

Mioni, G., Capizzi, M., and Stablum, F. (2020). Age-related changes in time production and reproduction tasks: involvement of attention and working memory processes. *Aging Neuropsychol. Cogn.* 27, 412–429. doi: 10.1080/13825585.2019.1626799

Norman, J. F., Cheeseman, J. R., Baxter, M. W., Thomason, K. E., Adkins, O. C., and Rogers, C. E. (2014). Aging and visual length discrimination: sequential dependencies, biases and the effects of multiple implicit standards. *Vis. Res.* 98, 89–98. doi: 10.1016/j.visres.2014.03.011

Perbal, S., Droit-volet, S., Isingrini, M. and Pouthas, V. (2002). Relationships between age-related changes in time estimation and age- related changes in processing speed, attention and memory. *Aging Neuropsychol. Cogn.* 9, 201–216.

Roach, N. W., McGraw, P. v., Whitaker, D. J., and Heron, J. (2017). Generalization of prior information for rapid Bayesian time estimation. *Proc. Nat. Acad. Sci. U S A* 114, 412–417. doi: 10.1073/pnas.1610706114

Sciutti, A., Burr, D., Saracco, A., Sandini, G., and Gori, M. (2014). Development of context dependency in human space perception. *Exp. Brain Res.* 232, 3965–3976. doi: 10.1007/s00221-014-4021-y

Scurry, A. N., Vercillo, T., Nicholson, A., Webster, M., and Jiang, F. (2019). Aging impairs temporal sensitivity, but not perceptual synchrony, across modalities. *Multisens. Res.* 32, 671–692. doi: 10.1163/22134808-20191343

Stocker, A. A., and Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nat. Neurosci.* 9, 578–585. doi: 10.1038/nn1669

Turgeon, M., Lustig, C., and Meck, W. H. (2016). Cognitive aging and time perception: roles of bayesian optimization and degeneracy. *Front. Aging Neurosci.* 8:102. doi: 10.3389/fnagi.2016.00102

Watson, A. B., and Pelli, D. G. (1983). QUEST: a bayesian adaptive psychometric method. *Percept. Psychophys.* 33, 113–120. doi: 10.3758/bf03202828

Weiss, Y., Simoncelli, E. P., and Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nat. Neurosci.* 5, 598–604. doi: 10.1038/nn0602-858

Wittmann, M., and Lehnhoff, S. (2005). Age effects in perception of time. *Psychol. Rep.* 97, 921–935. doi: 10.2466/pr0.97.3.921-935

frontiers | Frontiers in Human Neuroscience

TYPE Brief Research Report
PUBLISHED 30 November 2022
DOI 10.3389/fnhum.2022.1000598

OPEN ACCESS

EDITED BY
Elena Nava,
University of Milano-Bicocca, Italy

REVIEWED BY
Ilaria Berteletti,
Gallaudet University, United States
Lorenzo Vignali,
University of Trento, Italy

*CORRESPONDENCE
Margot Buyle
margot.buyle@uclouvain.be

SPECIALTY SECTION
This article was submitted to
Sensory Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 22 July 2022
ACCEPTED 07 November 2022
PUBLISHED 30 November 2022

CITATION
Buyle M and Crollen V (2022)
Deafness and early language
deprivation influence arithmetic
performances.
Front. Hum. Neurosci. 16:1000598.
doi: 10.3389/fnhum.2022.1000598

Deafness and early language deprivation influence arithmetic performances

Margot Buyle* and Virginie Crollen

Psychological Sciences Research Institute (IPSY) and Institute of Neuroscience (IoNS), Université catholique de Louvain, Louvain-la-Neuve, Belgium

It has been consistently reported that deaf individuals experience mathematical difficulties compared to their hearing peers. However, the idea that deafness and early language deprivation might differently affect verbal (i.e., multiplication) vs. visuospatial (i.e., subtraction) arithmetic performances is still under debate. In the present paper, three groups of 21 adults (i.e., deaf signers, hearing signers, and hearing controls) were therefore asked to perform, as fast and as accurately as possible, subtraction and multiplication operations. No significant group effect was found for accuracy performances. However, reaction time results demonstrated that the deaf group performed both arithmetic operations slower than the hearing groups. This group difference was even more pronounced for multiplication problems than for subtraction problems. Weaker language-based phonological representations for retrieving multiplication facts, and sensitivity to interference are two hypotheses discussed to explain the observed dissociation.

KEYWORDS

deafness, arithmetic, subtraction, multiplication, sign language, verbal, visuospatial

Introduction

A converging body of evidence in the numerical cognition field suggests that different arithmetic operations rely on distinct neuro-cognitive processes. Indeed, while subtraction is solved using visuospatial procedures (Dehaene, 1992; Campbell and Xue, 2001; Robinson, 2001; Seyler et al., 2003; Thevenot and Barrouillet, 2006; Barrouillet et al., 2008; Prado et al., 2014) and visuospatial shifts of attention (Li et al., 2018; Salvaggio et al., 2022) multiplication is, in contrast, rote learnt and stored in verbal memory (Verguts and Fias, 2005). Visuospatial skills accordingly predict subtraction, but not multiplication operations. Language skills inversely predict multiplication but not subtraction (Lee and Kang, 2002; Guez et al., 2022). At the neural level, subtraction has been linked to an increased activity of the parietal cortex, typically associated with quantity and visuospatial processing. Multiplication, on the other hand, relies on verbal brain areas of the left hemisphere (Lee, 2000; Zhou et al., 2007; Prado et al., 2011). In neuropsychology, impairments in phonological processing (e.g., dyslexic individuals) induce marked difficulties in multiplication fact retrieval but no impairment in subtraction (Simmons and Singleton, 2008; Boets and De Smedt, 2010; De Smedt and Boets, 2010). Double dissociations have moreover been reported with some patients selectively impaired in subtraction (Dehaene and Cohen, 1997; van Harskamp and Cipolotti, 2001) and others presenting the exact opposite pattern of performance: a selective impairment in multiplication fact retrieval and a preservation of their subtraction performances (Cohen and Dehaene, 2000; Cappelletti et al., 2001; van Harskamp and Cipolotti, 2001; Sandrini et al., 2003).

When taking the link that exists between language skills and arithmetic processing into account, it is not surprising to see that deaf individuals, who often experience some level of language deprivation in early childhood, present poorer numerical abilities than their hearing peers (see Buyle et al., 2021 for a review). A delay of 2 to 3.5 years on mathematical achievement tests (Nunes and Moreno, 2002; Bull et al., 2005) has indeed been highlighted and appears to be more pronounced in verbal numerical tasks (e.g., see Nunes et al., 2009 for multiplicative reasoning; Andin et al., 2014 for relational statements, Serrano Pau, 1995; Titus, 1995; Kelly et al., 2003 for fractions) than in visuospatial numerical tasks. In line with this, the absence of the SNARC effect in a verbal parity judgement task vs. the presence of the SNARC effect in a visuospatial number comparison task was recently shown in one of our previous studies (Buyle et al., 2022).

These observations were interestingly assumed to be caused by some linguistic aspects (Serrano Pau, 1995; Kelly and Mousley, 2001; Kelly et al., 2003; Pagliaro, 2010). In contrast to oral languages, sign languages are formed by several visual components such as the configuration, movement, orientation and location of the hands in space, the body posture, the facial expression and the movement of the mouth (Emmorey, 2002; Sandler and Lillo-Martin, 2006). These visual and motor aspects of sign language have already been shown to impact cognitive processes such as memory (Wilson and Emmorey, 1997) and reading (Quandt and Kubicek, 2018). Alpha and Beta EEG signals were for example found to be different when deaf signers read English words whose American Sign Language translations use two hands vs. one hand (Quandt and Kubicek, 2018). This result demonstrates the involvement of the sensorimotor system in cross-linguistic translation and supports the Dual-Route Cascade (DRC) model proposed by Elliott et al. (2011). This model suggests that the cognitive system involved in reading is fundamentally the same in deaf as in hearing (see the DRC model of Coltheart et al., 2001), but the types of activated units are different: visemes and phonemes for multimodal deaf bilingual vs. phonemes for monolingual or unimodal bilingual hearing individuals. L1 and L2 lexicons are both activated when deaf signers are reading. The viseme-phoneme translation that occurs in deaf signers can therefore affect their reading speed and proficiency. Associations between sign phonology and reading skills (Mayberry et al., 2011; Rudner et al., 2012) were accordingly reported in deaf individuals (Davis and Kelly, 2003).

As a close correlation between phonological awareness and arithmetic problem solving has also been repeatedly observed (De Smedt et al., 2010), the parallel between reading and arithmetic is tempting. The fact that sign languages use the entire body in a spatial-visual-somatic way may, for example, preserve or even positively impact (Chinello et al., 2012) the visuospatial arithmetic abilities of deaf individuals. In contrast, the fact that deaf signers do not easily access the phonology of verbal languages or access it through a viseme-phoneme translation may, in contrast, negatively impact their verbal arithmetic abilities (as already observed in reading, see Elliott et al., 2011). While this hypothesis is tempting, recent studies nevertheless failed to demonstrate clear results supporting this claim. While Andin et al. (2014) demonstrated that deaf signers perform worse on multiplication than on subtraction operations, more recent studies failed to demonstrate this dissociation (Andin et al., 2019). Mixed conclusions can also be found at the brain level. While an fMRI study showed that the right horizontal intraparietal sulcus was more activated in deaf signers as compared to hearing during multiplication operations (Andin et al., 2019), more recent studies (Andin et al., 2022; Berteletti et al., 2022) highlighted a comparable dissociation between the brain networks supporting multiplication and subtraction in deaf and hearing participants. There is therefore an urgent need to better characterise the impact deafness and its related language experience may have on arithmetic processing.

To do so, we will ask Belgian deaf signers, hearing signers and hearing controls to perform easy and difficult subtraction and multiplication operations. In Belgium, there are few options regarding education of deaf children. First, there exists the special-need education schools, but sign language is not provided as instruction language since the teachers are often hearing and using spoken language. Second, there is the regular school system with the presence of a sign language interpreter. However, the deaf child has to be confident with sign language before he/she can benefit from “translated” classes. A third and last option is the bilingual-bicultural education, which offers deaf children all the opportunities to get into contact with both spoken and signed languages, and both cultures. Unfortunately, not many schools provide this educational system in Belgium. Many Belgian deaf signers therefore consider sign language as their preferred communication method but were taught arithmetic in another spoken language. Including hearing signers in this study will therefore allow us to examine whether the arithmetic difficulties experienced by deaf signers are merely linked to the use of sign language or to the use of sign language as mother tongue (L1) while being taught arithmetic in a second spoken language (L2: Dutch or French). As several studies already demonstrated that unimodal bilingualism can impact number and arithmetic processing (Van Rinsveld et al., 2015, 2016, 2017; Lachelin et al., 2022), there is no reason to believe that number transcoding in bimodal bilinguals could not have any impact on arithmetic performances.

Finally, as recent behavioural (De Visscher and Noël, 2014a,b) and brain (De Visscher et al., 2018) findings on hearing people suggest that individual differences in multiplication fact knowledge may be partly due to differences in sensitivity to interference (De Visscher and Noël, 2013), we also decided to investigate this concept. It is based on the interference model of Campbell (1987) and Campbell (1995) according to which arithmetic facts involve various combinations of the digits 0 to 9, and therefore consist of very similar associations between two operands and the answer. As the similarity between the items to remember can cause memory interference (Oberauer and Lange, 2008), learning arithmetic facts that share a lot of common features can therefore be considered as highly interfering for the memory (Wickelgren, 1979). Individuals with higher sensitivity to interference therefore experience more proactive overlap from previously learned problems during arithmetic fact retrieval (De Visscher et al., 2018). A central executive impairment can therefore cause difficulties in arithmetic fact retrieval (Kaufmann, 2002; Temple and Sherwood, 2002; Noël et al., 2004; Barrouillet and Lépine, 2005), especially when a deficit in suppressing irrelevant information is present (i.e., inhibition) (Barrouillet et al., 1997; Pasolunghi et al., 1999; Censabella and Noël, 2004; Passolunghi and Siegel, 2004; Geary et al., 2012). De Visscher and Noël (2013) for example reported a case study of a dyscalculic individual showing hypersensitivity to interference in memory, and a circumscribed impairment to store arithmetic facts. Although deaf children and adults were often reported to present lower executive functioning than their hearing peers (Figueras et al., 2008; Hauser et al., 2008; Hintermair, 2013; Dye and Hauser, 2014; Hall et al., 2016; Botting et al., 2017; Jones et al., 2020), their sensitivity to interference while performing single-digit multiplication problems was never taken into account. This will be done in the present study.

To sum up, our study aims to investigate whether the arithmetic deficit in deaf individuals is: (1) global, or more specifically related to verbal numerical operations (i.e., multiplication problems); (2) linked to auditory deprivation, language deprivation, the mere use of sign language or the use of sign language as L1 while being taught arithmetic in a spoken L2 (Dutch or French that might not have been fully accessible despite the use of hearing aids); and (3) linked to the interference index of single-digit multiplication problems. If the arithmetic difficulties of deaf adults are global, their performance should be worse than the one of the hearing signers and hearing controls in both arithmetic operations. If deafness and its related language experience more strongly affects verbal operations, the difference between the deaf and the hearing adults should be bigger for the multiplication operations. Finally, as hearing signers were taught arithmetic in their mother tongue (i.e., French or Dutch), their later acquisition of sign language should not affect their arithmetic performances. They should therefore behave exactly as the hearing controls.

Methods

Participants

Three groups of 21 adults were recruited in the Dutch and French-speaking parts of Belgium: a group of congenitally deaf adults (12 females, 10 French, Mage = 39.1 years ± 2.92), a group of hearing signers (16 females, 11 French, Mage = 37.6 years ± 2.95), and a control group of hearing adults who did not know sign language (12 females, 10 French, Mage = 38.8 years ± 3.15) (see **Table 1** for a detailed description of the participants). All participants had normal or corrected-to-normal vision and no neurological problems. Hearing participants were matched to deaf participants for gender [X^2 (2, 63) = 2.19, $p = 0.33$], age [$F_{(2,60)} = 0.066$; $p = 0.94$, $\eta^2 = 0.002$], educational level [$F_{(2,58)} = 2.230$; $p = 0.12$, $\eta^2 = 0.071$], handedness [X^2 (2, 63) = 1.11, $p = 0.58$], and mother tongue (French vs. Dutch) [X^2 (2, 63) = 0.13, $p = 0.94$]. Hearing signers reported a minimum level of B1 (i.e., intermediate CEFR level) for sign language (see **Supplementary Table 1** for more details). Most (13) deaf individuals reported sign language as their mother tongue. Only seven deaf participants indicated being born in a deaf family, but six deaf participants indicated sign language as their mother tongue although not having any relatives with hearing problems. On the other hand, one deaf indicated having Dutch with gestures as mother tongue, and eight deaf participants reported acquiring sign language later in their life (2 to 20 years old), however, they were fluent in sign language and indicated it as their preferred way of communication (see **Supplementary Table 2** for more details). Both oral and written instructions in Dutch and in French were given, as well as instruction videos in sign language for deaf participants. Questions could be asked to the researcher, who is basic proficient in sign language. When really experiencing a language barrier, questions were answered in a written manner. Participants provided their written informed consent and the procedures were in line with the Declaration of Helsinki. The study was approved by the "Comité d'Ethique Hospitalo-Facultaire Saint-Luc-UCLouvain" (2019/19AOU/357).

TABLE 1 Characteristics of participants.

Subject	Age	Sex	Handedness	Onset	Cause	Formal school years (after primary school)
1	56	F	R	0	Hereditary	13
2	47	M	L	0	Rubella	6
3	26	F	R	3 years	Meningitis	12
4	26	F	R	0–1 year	Unknown	7
5	48	M	R	0	O_2 insufficiency	6
6	51	F	R	0	Meningitis	15
7	28	M	R	0	Genetic	14
8	50	M	L	0	Genetic	7
9	37	F	R	0	Genetic	12
10	49	F	R	0	Rubella	7
11	23	F	R	0	Unknown	9
12	43	M	R	0	Hereditary	5
13	24	F	R	0	Unknown	11
14	20	M	R	0	Unknown	7
15	53	M	R	0	Hereditary	6
16	53	F	R	0	Hereditary	6
17	35	M	R	0	Unknown	9
18	63	F	R	0	Hereditary	N/A
19	35	M	L	0	Genetic	8
20	37	F	R	0	Nerf atrophy	12
21	22	F	R	0	Unknown	10
22	21	F	R	0	CMV	8
23	23	F	R	/	/	9
24	30	F	R	/	/	12
25	23	F	R	/	/	11
26	31	F	R	/	/	N/A
27	58	F	R	/	/	12
28	23	F	R	/	/	7
29	26	F	R	/	/	12
30	29	F	R	/	/	9
31	51	F	R	/	/	17
32	32	F	R	/	/	11
33	41	F	R	/	/	9
34	23	F	R	/	/	8
35	29	F	L	/	/	15
36	57	F	L	/	/	14
37	28	F	R	/	/	16
38	30	F	R	/	/	11
39	50	M	R	/	/	10
40	63	M	R	/	/	9
41	51	M	R	/	/	10
42	54	M	R	/	/	6
43	38	M	R	/	/	12
44	55	F	R	/	/	11
45	23	M	R	/	/	12
46	23	M	R	/	/	9
47	23	F	R	/	/	10
48	20	F	R	/	/	8
49	38	F	R	/	/	12
50	50	M	R	/	/	11
51	57	F	R	/	/	6
52	46	M	R	/	/	6
53	66	M	R	/	/	11
54	31	M	R	/	/	16
55	57	M	R	/	/	10
56	50	F	R	/	/	9
57	39	M	R	/	/	8
58	38	F	R	/	/	9
59	25	F	R	/	/	14
60	36	F	L	/	/	11
61	49	M	R	/	/	7
62	47	F	R	/	/	7
63	20	F	R	/	/	9
64	21	F	R	/	/	10

R, right-handed; L, left-handed; F, female; M, male; CMV, cytomegalovirus.

Task and procedure

Participants had to solve two different arithmetic problems, namely subtraction problems and multiplication problems, which were presented on a computer screen in black font (Courier New font and size 42) on a grey background. Each category of operations consisted of 20 problems to solve: some easy operations (i.e., without carry-over for subtraction; one-digit number × one-digit number for multiplication) and some difficult operations (i.e., with carry-over for subtraction problems; two-digit number × one-digit number for multiplication problems; see **Table 2**). The participants first had to press the space bar when they knew the answer (to collect correct reaction times), and then use the keyboard to write down their answers. Operations were presented in a fixed order starting with subtraction problems and then multiplication problems. Easy operations were also presented before the difficult ones. This was done to not discourage deaf participants who are known to experience difficulties with arithmetic (Hyde et al., 2003; Kelly et al., 2003; Bull et al., 2011). Operations were presented on the screen until the participant pressed the space bar to be able to indicate their answer. The accuracy and reaction times of the responses were measured. Subjects executed the task in a silent room where the task was presented and the responses were recorded using E-Prime 2.0 software running on a Dell computer with Windows XP as operating system.

Statistical analysis

Statistical analyses were carried out using IBM SPSS statistics 26 software for Mac OS Monterey 12.0.1 (Armonk, NY, USA). Statistical significance was set at $p < 0.05$

TABLE 2 Operations presented to the participants.

Subtraction	Carryover	Level	Multiplication	Format	Level	Interference index
6–2	No	Easy	3 × 2	U × U	Easy	0
8–5	No	Easy	4 × 3	U × U	Easy	10
7–3	No	Easy	5 × 4	U × U	Easy	8
9–4	No	Easy	6 × 5	U × U	Easy	6
17–5	No	Easy	8 × 6	U × U	Easy	11
27–4	No	Easy	2 × 7	U × U	Easy	4
48–6	No	Easy	9 × 4	U × U	Easy	9
54–3	No	Easy	4 × 8	U × U	Easy	25
63–9	Yes	Difficult	7 × 9	U × U	Easy	17
35–6	Yes	Difficult	5 × 7	U × U	Easy	7
21–7	Yes	Difficult	13 × 5	DU × U	Difficult	N/A
44–8	Yes	Difficult	24 × 4	DU × U	Difficult	N/A
24–11	No	Easy	38 × 3	DU × U	Difficult	N/A
58–33	No	Easy	17 × 6	DU × U	Difficult	N/A
27–15	No	Easy	56 × 2	DU × U	Difficult	N/A
47–22	No	Easy	61 × 3	DU × U	Difficult	N/A
52–39	Yes	Difficult	72 × 2	DU × U	Difficult	N/A
43–27	Yes	Difficult	29 × 5	DU × U	Difficult	N/A
65–39	Yes	Difficult	45 × 4	DU × U	Difficult	N/A
54–18	Yes	Difficult	31 × 6	DU × U	Difficult	N/A

U, one-digit number; DU, two-digits number.

for all computations. Data were checked for normality of distribution and presented as Mean ± Standard Error (SE). Accuracy scores and reaction times (in ms) were measured. A binary Generalised Linear Mixed Model (GLMM) was run on accuracy scores (correct or not correct). A GLMM was run on the reaction time data (only reactions times for correct answers were included), indicating gamma distribution. One random factor was included in all analyses because of its significant contribution to the variance (i.e., subjects). The fixed factors included *Group* (deaf, hearing signers, hearing controls), *Operation* (subtraction, multiplication) and *Level* (easy, difficult), as well as their interactions. Given that deaf adults often experience executive functioning difficulties due to language deprivation, we hypothesised that our deaf group might be more sensitive to interference. To investigate if deaf individuals are indeed more affected by the interference index of single-digit multiplication problems, we performed a GLMM indicating gamma distribution with *Group*, *Interference index* and its interaction as fixed factors, and reaction time as dependent variable. Interference indexes were taken from De Visscher and Noël (2014b), since the authors calculated the interference index for all the 36 single-digit multiplication problems. As this index is not prone to change, and always remains the same for one specific operation, we could use this value directly in our analysis (see **Figure 1** of De Visscher and Noël, 2014b; **Table 2** for the related interference index of the single-digit multiplications presented in this study). Sequential Bonferroni adjusted significance level was applied when appropriate. Only the first and last model of the GLMM analyses where all non-significant interactions were not considered anymore in the model are reported, to (1) obtain a model that is quite easy to interpret, and (2) to gain power for the remaining parameters to detect significance. Outlier data were removed from statistical analysis when 3 standard deviations out of the mean (i.e., one deaf participant was removed from the testing sample). For subtraction problems, 0.95% outlier data were removed for accuracy and for reaction times in the deaf group; 0.48% for accuracy and 2.38% for reaction times in the hearing signer group, and 1.90% of the reaction times for the hearing controls group. For multiplication, the proportion of outliers for accuracy was 1.20% for the deaf group, 0.24% for the hearing signers, and 0.48% for the hearing controls. A total of 2.38% was removed of the reaction times for the hearing signer as well as for the hearing control group, and 2.62% for the deaf group.

Results

Accuracy

A binary GLMM was run on the accuracy scores as described above. No main effect of *Operation* [$F_{(1,2494)} = 2.55$; $p = 0.11$] or *Group* [$F_{(2,2494)} = 0.13$; $p = 0.88$] was shown.

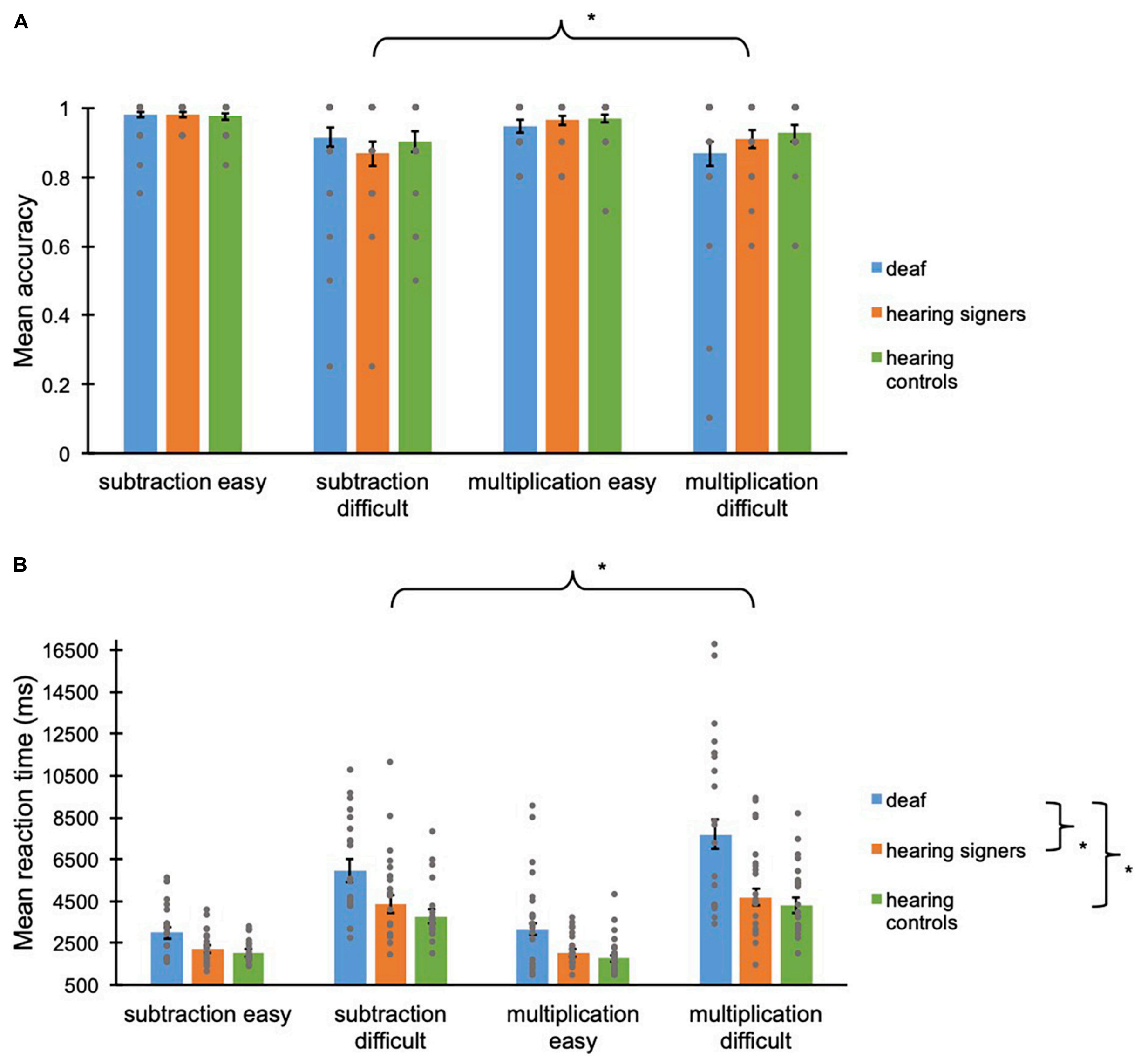


FIGURE 1
(A) Mean accuracy scores (proportions) and **(B)** mean reaction times (ms) for deaf (in blue), hearing signers (in orange), and hearing controls (in green) for the two different operations, and the two different levels of the arithmetic task. Error bars represent the standard error of the means. Asterisks represent significant difference. Grey points represent individual mean scores.

Nevertheless, a significant main effect of *Level* [$F_{(1,2494)} = 57.8$; $p < 0.001$] and a significant *Level x Operation* interaction was observed [$F_{(1,2494)} = 4.42$; $p = 0.036$]. The *Group* × *Operation* [$F_{(2,2494)} = 2.36$; $p = 0.095$], *Group* × *Level* [$F_{(2,2494)} = 0.27$; $p = 0.77$], and *Group* × *Operation* × *Level* [$F_{(2,2494)} = 0.23$; $p = 0.80$] interactions were not significant. The final GLMM was run with the only significant interaction included and led to the same conclusion: Significantly higher accuracy scores were observed for the easy operations ($m = 0.97$, $se = 0.005$) compared to the difficult operations ($m = 0.90$, $se = 0.015$, $p < 0.001$). A significant difference between subtraction problems and multiplication problems was found for the easy operations only, where the accuracy scores of multiplication problems were lower ($m = 0.96$, $se = 0.009$) than those for subtraction problems ($m = 0.98$, $se = 0.005$, $p = 0.025$) (see **Figure 1A**).

Reaction times

Regarding the reaction times, the GLMM indicated a significant difference for *Operation* [$F_{(1,2265)} = 4.26$; $p = 0.039$], *Group* [$F_{(2,\ 2265)} = 10.1$; $p < 0.001$], and *Level* [$F_{(1,2265)} = 938$; $p < 0.001$]. No significant *Group* × *Level* interaction [$F_{(2,2265)} = 0.32$; $p = 0.72$] and no *Group* × *Operation* × *Level*

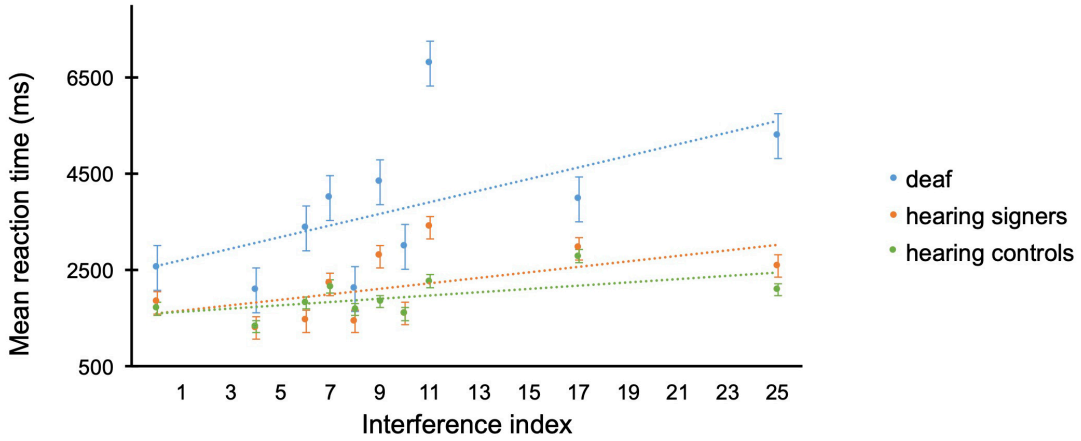


FIGURE 2
Mean reaction times (ms) per interference index for the easy multiplication operations in the different groups (deaf in blue, hearing signers in orange, hearing controls in green) of the arithmetic task. Error bars represent the standard error of the means.

interaction [$F_{(2,2265)}$ = 0.51; p = 0.60] was seen. However, significant *Group* × *Operation* [$F_{(2,2265)}$ = 4.00; p = 0.018] and *Operation* × *Level* [$F_{(1,2265)}$ = 16.4; p < 0.001] interactions were observed. The final GLMM was run including the two significant interactions and led to the same conclusion: Subtraction problems (m = 3318, se = 162) were solved faster than multiplication problems (m = 3492, se = 170, p = 0.040). Deaf adults (m = 4551, se = 372) were slower than hearing signer adults (m = 3114, se = 254, p = 0.003), and hearing control adults (m = 2783, se = 227, p < 0.001). Hearing signers did not perform differently compared to hearing controls (p = 0.33). Responses to difficult operations were slower (m = 4987, se = 246) than responses to easy operations (m = 2323, se = 112, p < 0.001). The difference between deaf (m = 4917, se = 417 *for multiplication problems and* m = 4213, se = 355 *for subtraction problems*) and hearing signers (m = 3122, se = 263 *for multiplication problems*, p = 0.001 and m = 3106, se = 262 *for subtraction problems*, p = 0.024) as well as between deaf and hearing controls (m = 2775, se = 234 *for multiplication problems*, p < 0.001 and m = 2791, se = 236 *for subtraction problems*, p = 0.003) was bigger for the multiplication problems than for the subtraction problems. The difference between subtraction problems and multiplication problems was only found for the difficult operations (m = 5379, se = 280 *for multiplication problems and* m = 4623, se = 248 *for subtraction problems*, p < 0.001), and not for the easy ones (m = 2267, se = 117 *for multiplication problems and* m = 2381, se = 120 *for subtraction problems*, p = 0.13) (see **Figure 1B**). Similar results are found when including years of formal education as covariate (see **Supplementary material**). Moreover, no speed accuracy trade-off was observed in any groups and/or any operations.

When investigating the interference index using a GLMM, a main effect of *Group* [$F_{(2,561)}$ = 7.96; p < 0.001], and *Interference index* [$F_{(9,561)}$ = 23.0; p < 0.001] was observed together with a significant *Group* × *Interference index* interaction [$F_{(18,561)}$ = 2.10; p = 0.005]. Deaf (m = 3010, se = 309) performed slower than hearing signers (m = 1975, se = 202), p = 0.011, and hearing controls (m = 1727, se = 177), p = 0.001 (see **Figure 2**). *Post-hoc* analyses on the *Group* × *Interference index* interaction can be found as **Supplementary Table 3**. In general, a pattern indicating more significant group differences with augmenting interference index was observed.

Discussion

Deafness has been indicated as a risk factor for mathematical difficulties, where the differences between signed and spoken language, less exposure to numerical language, and differences in domain-general processing are suggested to contribute mostly to this phenomenon (see Santos and Cordes, 2022 for a review). The challenges that deaf individuals experience with mathematical abilities have indeed been consistently demonstrated in the literature over the last decades (e.g., Wollman, 1965; Hine, 1970; Wood et al., 1986; Bull, 2008), and are thought to primarily lie in the acquisition of verbal number concepts such as counting, fractions, and, more importantly for our purposes, arithmetic skills (Titus, 1995; Leybaert and Van Cutsem, 2002; Kritzer, 2009; Pagliaro and Kritzer, 2013). While the underperformance of deaf individuals in arithmetic has been highlighted by different mathematical assessment tests (Bull et al., 2011 for math achievement test; Hyde et al., 2003 for arithmetic word problems; Kelly et al., 2003 for relational statements; Pagliaro and Ansell, 2012 for arithmetic story problems), the differential impact deafness and language deprivation may have on verbal vs. visuospatial arithmetic operations is less clear. To examine this possible dissociation, deaf signer, hearing signer and hearing control adults were asked to solve easy vs. difficult subtraction and multiplication operations.

Overall, our results demonstrated that performances were lower for the difficult operations as compared to the easy ones, and lower for multiplication problems as compared to subtraction problems in all three groups. This accuracy difference between subtraction problems and multiplication problems was, however, only present for the easy operations, while the reaction time difference between the same operations was only present for the difficult operations. But, most importantly for our purposes, and in contrast to Andin et al. (2019), we managed to highlight a difference between deaf signers and hearing adults at the behavioural level. Group differences were found for reaction times–but not for accuracy scores–(i.e., the deaf were slower than the two hearing groups), and these group differences were larger for multiplication problems than for subtraction problems. The discrepancy between our study and the one of Andin et al. (2019) probably comes from the fact that different groups of participants and different tasks were tested in these two studies. Indeed, in Andin et al. (2019), the deaf participants group only included native signers. Participants were moreover required to verify (and not calculate) the results of subtraction and multiplication problems. In a verification task, individuals can decide that the answer is false on the basis of plausibility judgements (e.g., Duverne and Lemaire, 2004, 2005; Hinault et al., 2015). Solution times are therefore not representative of the genuine time it takes to solve an arithmetic operation in an ecological situation. It is finally worth mentioning that only single-digit operations were included in this study. This level of arithmetic reasoning might have not been sufficient enough to highlight group differences in adults (Andin et al., 2022).

Recent years have seen a surge in empirical studies examining the role of language in accounting for cross-language disparities in children's number understanding and arithmetic competence (Fuson and Kwon, 1992; Rasmussen et al., 2006; Wang et al., 2008; Krinzinger et al., 2011; Göbel et al., 2014). It has for example been suggested that the superior arithmetic performance of Chinese and other Asian students could be explained by the relative *linguistic transparency* of the Asian counting systems (Fuson and Kwon, 1992; Miller et al., 2005), which gives a clear and consistent representation of the base-ten system (contrarily to the base-five system of the sign languages used in Belgium). In line with this, when considering bilingual individuals, the language in which arithmetic was learned seems to have a remaining advantage on performance. Van Rinsveld et al. (2016), for example, found better performances on arithmetic problem solving in German than in French, since German is the first learned language at the Luxembourgish school system. While comparisons across different auditory languages have been made, the present study aimed to examine the impact of sign language use on arithmetic problem solving. Since the obtained results indicated no significant difference between hearing signer and hearing control adults, one could assume that it is rather the usage of sign language as L1, while having learned multiplication in spoken language, that influences multiplication performances and not the knowledge of sign language *per se*. Belgian deaf signers could possibly use a visuospatial route while solving multiplication operations. Hearing individuals would in contrast directly access the verbal route. The visuospatial detour that deaf signers experience could explain why solving multiplication operations requires them more time and resources (i.e., cognitive load). Hearing individuals may not prevent themselves from relying on the phonological aspects of the presented stimuli, while deaf signers may experience some issues in accessing the verbal associations of multiplication facts. This hypothesis is, however, speculative and should be further tested in the future.

As the control groups (hearing controls and hearing signers) have experienced typical language development with typical language access from birth, they differ from the deaf group in language modality and in hearing status but also in early language access. This delay in accessing language can therefore be the main factor subtending the arithmetic difficulties of our deaf sample. Signed languages are indeed complete, natural languages that consist of their own unique visual grammar and syntax (Stokoe et al., 1965). Consistently with the fact that typically developing children with higher phonological awareness are better in forming verbal representations of multiplicative relations between two numbers (De Smedt et al., 2010; Berteletti et al., 2014), deaf children born to deaf parents who are fluent signers, do not display the same difficulties with mathematics as those with language deprivation early in development (e.g., Kritzer, 2009; Mousley and Kurz, 2015; Hrastinski and Wilbur, 2016). This distinction highlights an important relationship between language access and acquiring numerical concepts, or the importance of mastering sign language phonology to perform well on multiplication in the deaf signers population (e.g., Berteletti et al., 2022). **Supplementary Figure 1** representing individual data indicates that the early deaf signers seem to be more efficient than the later deaf signers of our sample. Late deaf signers probably experienced some early language deprivation and possibly limited access to spoken languages during the critical years for learning mathematics. Speculatively, if all opted to acquire sign language later in life (i.e., after the age of 3 years/o), it is probably because the quality of the auditory input or the difficulty in processing it was non-negligible (see **Supplementary Figure 1**). More systematically comparing native or late signers to early or late cochlear implanted deaf individuals would definitely help to understand whether the mathematical difficulties deaf often experience originate from auditory deprivation *per se* or from a delay in accessing and mastering verbal or visual languages.

While Andin et al. (2019) failed to find behavioural group differences in reaction time and accuracy on their arithmetic task, they nevertheless highlighted differences in the neural networks deaf signers and hearing non-signers engage to calculate (but see Andin et al., 2022; Berteletti et al., 2022). Whereas language related brain regions in the left cerebral hemisphere are usually recruited for arithmetic fact retrieval (Dehaene et al., 2003), stronger activation of the right horizontal intraparietal sulcus was found in deaf signers compared to hearing non-signers. This indicates that deaf signers may solve multiplication operations by relying on magnitude manipulation to a larger extent than their hearing peers (Andin et al., 2019). They could therefore be more sensitive to the numerical magnitude of the operations presented (see **Supplementary Figures 2, 3**). In line with this, we have to admit that our multiplication problems tend to be larger in overall magnitude than our subtraction problems. The greatest difficulty that we observe in deaf signers for the multiplication problems may therefore lie in their greater quantity processing rather than to the visuo-phonemes translation they require. If small multiplication problems (Siegler, 1988) are solved by direct memory retrieval, it is true that larger multiplication problems are more likely to be split up in easier problems and then involve visuospatial procedures to manipulate intermediate calculations and the magnitude of the final result (LeFevre et al., 1996; Thevenot et al., 2001, 2007; Núñez-Peña et al., 2011). As splitting up the operation in easier problems involves retrieving them as arithmetic facts, we are nevertheless convinced that multiplication problems (easy and difficult) require more language and memory processes than subtraction operations. Retrieving arithmetic facts and manipulating intermediate calculations could therefore be difficult for deaf individuals. This makes even more sense if we consider that: (1) language deprivation correlates with executive functioning difficulties (Hall et al., 2016; Botting et al., 2017; Jones et al., 2020; Ribner et al., 2022); and (2) spoken language is temporal and has been shown to lead to higher serial spans than signed information in serial recall tasks (Bavelier et al., 2008). As deaf signers who present language deprivation were shown to perform significantly poorer on executive functioning tasks than hearing individuals (Figueras et al., 2008; Hauser et al., 2008; Hintermair, 2013; Dye and Hauser, 2014; Hall et al., 2016; Botting et al., 2017; Jones et al., 2020), they could therefore heighten more sensitivity to interference. This assumption was supported by our interference index analysis. Moreover, as temporal order is maintained more efficiently in auditory-based representations than in visually-based representations (Paivio and Csapo, 1971; Watkins and Watkins, 1980; Watkins et al., 1992), speakers would rely more on temporal encoding, while signers would rely more on spatial encoding (Wilson, 2001). Deaf individuals could therefore experience more problems to solve multiplication operations, as they might be less efficient to learn a sequence of multiplication facts.

To conclude, our findings are in line with several previous studies suggesting that deaf individuals have no deficits in their numerical representation of magnitude information (i.e., similar accuracy scores), but might experience a less efficient processing (i.e., slower reaction times) of basic numerical information (Epstein et al., 1994; Iversen et al., 2004; Bull et al., 2005; Chinello et al., 2012; Rodriguez-Santos et al., 2014). This less efficient processing is, in our case, more pronounced for multiplication than for subtraction operations and could be explained by several mutually not exclusive reasons: (1) the fact that deaf individuals have delayed and therefore less automatic access than their hearing peers to the verbal phonological loop (Elliott et al., 2011); (2) the fact that deaf individuals might show higher sensitivity to the magnitude of the arithmetic operation presented; and (3) the fact that deaf individuals might show higher sensitivity to interference for multiplication operations (De Visscher and Noël, 2013, 2014a,b; De Visscher et al., 2018).

Conclusion

This study investigated how deafness and its related variable language experience, including language deprivation, shapes verbal vs. visuospatial arithmetic performances. Although the accuracy scores between deaf signers, hearing signers and hearing controls did not differ, the deaf signers showed significantly slower reaction times compared to the two hearing groups. Importantly, this significant group difference was larger for multiplication operations than for subtraction operations. These findings support the idea that numerical tasks relying on verbal processes are more strongly impacted by deafness and its following language experience, compared to numerical tasks implying visuospatial processes (Buyle et al., 2022). Further studies are, however, needed to better understand the mechanisms underlying this dissociation. Performances of deaf and hearing children should for example be compared on many more "easy" operations. Varying much more the magnitude of the operands and the interference index of the multiplications presented will help to understand the impact of these two factors on the arithmetic development of deaf individuals. Asking deaf and hearing children to perform multiplication operations under verbal vs. visuospatial load may similarly help to understand whether deaf and hearing signers use different processes to solve these operations (verbal processes in hearing vs. visuospatial processes or magnitude manipulation in deaf signers).

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, upon request to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by the "Comité d'Ethique Hospitalo-Facultaire Saint-Luc-UCLouvain" (2019/19AOU/357). The patients/participants provided their written informed consent to participate in this study.

Author contributions

MB: conceptualisation, methodology, software, formal analysis, investigation, data curation, visualisation, writing – original draft and review, and editing. VC: conceptualisation, methodology, software, writing – review and editing, supervision, validation, project administration, and funding acquisition. Both authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Incentive Grant for Scientific Research (MIS) granted by the Fund for Scientific Research–National Fund for Scientific Research (FRS–FNRS, Belgium) to VC (Grant no: F.4505.19).

Acknowledgments

We would like to thank all our participants. This research has benefitted from the statistical consult with Statistical Methodology and Computing Service, technological platform at UCLouvain–SMCS/LIDAM, UCLouvain.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2022.1000598/full#supplementary-material

References

Andin, J., Fransson, P., Dahlström, Ö, Rönnberg, J., and Rudner, M. (2019). The neural basis of arithmetic and phonology in deaf signing individuals. *Lang. Cogn. Neurosci.* 34, 813–825.

Andin, J., Rönnberg, J., and Rudner, M. (2014). Deaf signers use phonology to do arithmetic. *Learn. Individ. Differ.* 32, 246–253.

Andin, J., Elwér, Å, and Mäki-Torkko, E. (2022). Arithmetic in the adult deaf signing brain. *J. Neurosci. Res.* 98, 643–654. doi: 10.1002/jnr.24569

Barrouillet, P., Fayol, M., and Lathulière, E. (1997). Selecting between competitors in multiplication tasks: An Explanation of the errors produced by adolescents with learning difficulties. *Int. J. Behav. Dev.* 21, 253–275.

Barrouillet, P., and Lépine, R. (2005). Working memory and children's use of retrieval to solve addition problems. *J. Exp. Child Psychol.* 91, 183–204. doi: 10.1016/j.jecp.2005.03.002

Barrouillet, P., Mignon, M., and Thevenot, C. (2008). Strategies in subtraction problem solving in children. *J. Exp. Child Psychol.* 99, 233–251. doi: 10.1016/j.jecp.2007.12.001

Bavelier, D., Newport, E. L., Hall, M., Supalla, T., and Boutla, M. (2008). Ordered short-term memory differs in signers and speakers: Implications for models of short-term memory. *Cognition* 107, 433–459. doi: 10.1016/j.cognition.2007.10.012

Berteletti, I., Kimbley, S. E., Sullivan, S. J., Quandt, L. C., and Miyakoshi, M. (2022). Different language modalities yet similar cognitive processes in arithmetic fact retrieval. *Brain Sci.* 12:145.

Berteletti, I., Prado, J., and Booth, J. R. (2014). Children with mathematical learning disability fail in recruiting verbal and numerical brain regions when solving simple multiplication problems. *Cortex* 57, 143–155. doi: 10.1016/j.cortex.2014.04.001

Boets, B., and De Smedt, B. (2010). Single-digit arithmetic in children with dyslexia. *Dyslexia* 16, 183–191.

Botting, N., Jones, A., Marshall, C., Denmark, T., Atkinson, J., and Morgan, G. (2017). Nonverbal executive function is mediated by language: A study of deaf and hearing children. *Child Dev.* 88, 1689–1700. doi: 10.1111/cdev.12659

Bull, R. (2008). "Deafness, numerical cognition, and mathematics," in *Deaf Cognition: Foundations and Outcomes*, ed. P. H. M. Marschark (New York, NY: Oxford).

Bull, R., Marschark, M., and Blatto-Vallee, G. (2005). SNARC hunting: Examining number representation in deaf students. *Learn. Individ. Differ.* 15, 223–236.

Bull, R., Marschark, M., Sapere, P., Davidson, W. A., Murphy, D., and Nordmann, E. (2011). Numerical estimation in deaf and hearing adults. *Learn. Individ. Differ.* 21, 453–457. doi: 10.1016/j.lindif.2011.02.001

Buyle, M., Marlair, C., and Crollen, V. (2021). "Diversity dimensions in mathematics and language learning," in *Blindness and Deafness: A Window to Study the Visual and Verbal Basis of the Number Sense*, eds F. Annemarie, G. Erkan, and H. Moritz (Berlin: De Gruyter), 265–288.

Buyle, M., Vencato, V., and Crollen, V. (2022). Impact of deafness on numerical tasks implying visuospatial and verbal processes. *Sci. Rep.* 12:11150. doi: 10.1038/s41598-022-14728-3

Campbell, J. I., and Xue, Q. (2001). Cognitive arithmetic across cultures. *J. Exp. Psychol.* 130, 299–315.

Campbell, J. I. D. (1987). Network interference and mental multiplication. *J. Exp. Psychol.* 13, 109–123. doi: 10.1037/0278-7393.13.1.109

Campbell, J. I. D. (1995). Mechanisms of simple addition and multiplication: A modified network-interference theory and simulation. *Math. Cogn.* 1, 121–164.

Cappelletti, M., Butterworth, B., and Kopelman, M. (2001). Spared numerical abilities in a case of semantic dementia. *Neuropsychologia* 39, 1224–1239. doi: 10.1016/s0028-3932(01)00035-5

Censabella, S., and Noël, M. P. (2004). Interference in arithmetic facts: Are active suppression processes involved when performing simple mental arithmetic? *Curr. Psychol. Cogn.* 22, 635–671.

Chinello, A., de Hevia, M. D., Geraci, C., and Girelli, L. (2012). Finding the spatial-numerical association of response codes (SNARC) in signed numbers: Notational effects in accessing number representation. *Funct. Neurol.* 27, 177–185.

Cohen, L., and Dehaene, S. (2000). Calculating without reading: Unsuspected residual abilities in pure alexia. *Cogn. Neuropsychol.* 17, 563–583. doi: 10.1080/02643290050110656

Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychol. Rev.* 108, 204–256. doi: 10.1037/0033-295x.108.1.204

Davis, S. M., and Kelly, R. R. (2003). Comparing deaf and hearing college students' mental arithmetic calculations under two interference conditions. *Am. Ann. Deaf* 148, 213–221. doi: 10.1353/aad.2003.0018

De Smedt, B., and Boets, B. (2010). Phonological processing and arithmetic fact retrieval: Evidence from developmental dyslexia. *Neuropsychologia* 48, 3973–3981.

De Smedt, B., Taylor, J., Archibald, L., and Ansari, D. (2010). How is phonological processing related to individual differences in children's arithmetic skills? *Dev. Sci.* 13, 508–520. doi: 10.1111/j.1467-7687.2009.00897.x

De Visscher, A., and Noël, M.-P. (2013). A case study of arithmetic facts dyscalculia caused by a hypersensitivity-to-interference in memory. *Cortex* 49, 50–70. doi: 10.1016/j.cortex.2012.01.003

De Visscher, A., and Noël, M. P. (2014a). Arithmetic facts storage deficit: The hypersensitivity-to-interference in memory hypothesis. *Dev. Sci.* 17, 434–442. doi: 10.1111/desc.12135

De Visscher, A., and Noël, M. P. (2014b). The detrimental effect of interference in multiplication facts storing: Typical development and individual differences. *J. Exp. Psychol. Gen.* 143, 2380–2400. doi: 10.1037/xge0000029

De Visscher, A., Vogel, S. E., Reishofer, G., Hassler, E., Koschutnig, K., De Smedt, B., et al. (2018). Interference and problem size effect in multiplication fact solving: Individual differences in brain activations and arithmetic performance. *NeuroImage* 172, 718–727. doi: 10.1016/j.neuroimage.2018.01.060

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition* 44, 1–42. doi: 10.1016/0010-0277(92)90049-n

Dehaene, S., and Cohen, L. (1997). Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex* 33, 219–250.

Dehaene, S., Piazza, M., Pinel, P., and Cohen, L. (2003). Three parietal circuits for number processing. *Cogn. Neuropsychol.* 20, 487–506. doi: 10.1080/02643290244000239

Duverne, S., and Lemaire, P. (2004). Age-related differences in arithmetic problem-verification strategies. *J. Gerontol. Series B* 59, 135–142.

Duverne, S., and Lemaire, P. (2005). Arithmetic split effects reflect strategy selection: An adult age comparative study in addition comparison and verification tasks. *Can. J. Exp. Psychol.* 59, 262–278. doi: 10.1037/h0087479

Dye, M. W. G., and Hauser, P. C. (2014). Sustained attention, selective attention and cognitive control in deaf and hearing children. *Hear. Res.* 309, 94–102. doi: 10.1016/j.heares.2013.12.001

Elliott, E. A., Braun, M., Kuhlmann, M., and Jacobs, A. M. (2011). A dual-route cascaded model of reading by deaf adults: Evidence for grapheme to viseme conversion. *J. Deaf Stud. Deaf Educ.* 17, 227–243. doi: 10.1093/deafed/enr047

Emmorey, K. (2002). "Insights from sign language research," in *Language, Cognition & the Brain*, ed. N. Mahwah (Hillsdale, NJ: Lawrence Erlbaum Associates).

Epstein, K. I., Hillegeist, E. G., and Grafman, J. (1994). Number processing in deaf college students. *Am. Ann. Deaf* 139, 336–347. doi: 10.1353/aad.2012.0321

Figueras, B., Edwards, L., and Langdon, D. (2008). Executive function and language in deaf children. *J. Deaf Stud. Deaf Educ.* 13, 362–377. doi: 10.1093/deafed/enm067

Fuson, K. C., and Kwon, Y. (1992). Korean children's understanding of multidigit addition and subtraction. *Child Dev.* 63, 491–506. doi: 10.1111/j.1467-8624.1992.tb01642.x

Geary, D. C., Hoard, M. K., and Bailey, D. H. (2012). Fact retrieval deficits in low achieving children and children with mathematical learning disability. *J. Learn. Disabil.* 45, 291–307. doi: 10.1177/0022219410392046

Göbel, S. M., Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H.-C. (2014). Language affects symbolic arithmetic in children: The case of number word inversion. *J. Exp. Child Psychol.* 119, 17–25. doi: 10.1016/j.jecp.2013.10.001

Guez, A., Piazza, M., Pinheiro-Chagas, P., Peyre, H., Heude, B., and Ramus, F. (2022). Preschool language and visuospatial skills respectively predict multiplication and addition/subtraction skills in middle school children. *Dev. Sci.* 2022:e13316. doi: 10.1111/desc.13316

Hall, M. L., Eigsti, I.-M., Bortfeld, H., and Lillo-Martin, D. (2016). Auditory deprivation does not impair executive function, but language deprivation might: Evidence from a parent-report measure in deaf native signing children. *J. Deaf Stud. Deaf Educ.* 22, 9–21. doi: 10.1093/deafed/enw054

Hauser, P. C., Lukomski, J., and Hillman, T. (2008). "Development of deaf and hard-of-hearing students' executive function," in *Deaf Cognition: Foundations and Outcomes*, eds M. Marschark and P. C. Hauser (New York, NY: Oxford University Press), 286–308.

Hinault, T., Tiberghien, K., and Lemaire, P. (2015). Age-related differences in plausibility-checking strategies during arithmetic problem verification tasks. *J. Gerontol.* 71, 613–621. doi: 10.1093/geronb/gbu178

Hine, W. (1970). Attainments of children with partial hearing. *Teach. Deaf* 68, 129–135.

Hintermair, M. (2013). Executive functions and behavioral problems in deaf and hard-of-hearing students at general and special schools. *J. Deaf. Stud. Deaf. Educ.* 18, 344–359. doi: 10.1093/deafed/ent003

Hrastinski, I., and Wilbur, R. B. (2016). Academic achievement of deaf and hard-of-hearing students in an asl/english bilingual program. *J. Deaf Stud. Deaf Educ.* 21, 156–170. doi: 10.1093/deafed/env072

Hyde, M., Zevenbergen, R., and Power, D. (2003). Deaf and hard of hearing students' performance on arithmetic word problems. *Am. Ann. Deaf* 148, 56–64. doi: 10.1353/aad.2003.0003

Iversen, W., Nuerk, H. C., and Willmes, K. (2004). Do signers think differently? The processing of number parity in deaf participants. *Cortex* 40, 176–178. doi: 10.1016/s0010-9452(08)70940-7

Jones, A., Atkinson, J., Marshall, C., Botting, N., St Clair, M. C., and Morgan, G. (2020). Expressive vocabulary predicts nonverbal executive function: A 2-year longitudinal study of deaf and hearing children. *Child Dev.* 91, e400–e414. doi: 10.1111/cdev.13226

Kaufmann, L. (2002). More evidence for the role of the central executive in retrieving arithmetic facts – a case study of severe developmental dyscalculia. *J. Clin. Exp. Neuropsychol.* 24, 302–310. doi: 10.1076/jcen.24.3.302.976

Kelly, R. R., Lang, H. G., Mousley, K., and Davis, S. M. (2003). Deaf college students' comprehension of relational language in arithmetic compare problems. *J. Deaf. Stud. Deaf. Educ.* 8, 120–132. doi: 10.1093/deafed/eng006

Kelly, R. R., and Mousley, K. (2001). Solving word problems: More than reading issues for deaf students. *Am. Ann. Deaf* 146, 251–262. doi: 10.1353/aad.2012.0088

Krinzinger, H., Gregoire, J., Desoete, A., Kaufmann, L., Nuerk, H.-C., and Willmes, K. (2011). Differential language effects on numerical skills in second grade. *J. Cross Cult. Psychol.* 42, 614–629. doi: 10.1177/0022022111406252

Kritzer, K. L. (2009). Barely started and already left behind: A descriptive analysis of the mathematics ability demonstrated by young deaf children. *J. Deaf Stud. Deaf Educ.* 14, 409–421. doi: 10.1093/deafed/enp015

Lachelin, R., van Rinsveld, A., Poncin, A., and Schiltz, C. (2022). Number transcoding in bilinguals-A transversal developmental study. *PLoS One* 17:e0273391. doi: 10.1371/journal.pone.0273391

Lee, K., and Kang, S. (2002). Arithmetic operation and working memory: Differential suppression in dual tasks. *Cognition* 83, B63–B68. doi: 10.1016/s0010-0277(02)00010-0

Lee, K. M. (2000). Cortical areas differentially involved in multiplication and subtraction: A functional magnetic resonance imaging study and correlation with a case of selective acalculia. *Ann. Neurol.* 48, 657–661.

LeFevre, J.-A., Sadesky, G. S., and Bisanz, J. (1996). Selection of procedures in mental addition: Reassessing the problem size effect in adults. *J. Exp. Psychol.* 22, 216–230. doi: 10.1037/0278-7393.22.1.216

Leybaert, J., and Van Cutsem, M. N. (2002). Counting in sign language. *J. Exp. Child Psychol.* 81, 482–501. doi: 10.1006/jecp.2002.2660

Li, M., Liu, D., Li, M., Dong, W., Huang, Y., and Chen, Q. (2018). Addition and subtraction but not multiplication and division cause shifts of spatial attention. *Front. Hum. Neurosci.* 12:183. doi: 10.3389/fnhum.2018.00183

Mayberry, R. I., del Giudice, A. A., and Lieberman, A. M. (2011). Reading achievement in relation to phonological coding and awareness in deaf readers: A meta-analysis. *J. Deaf Stud. Deaf Educ.* 16, 164–188. doi: 10.1093/deafed/enq049

Miller, K. F., Kelly, M., and Zhou, Z. (2005). "Learning mathematics in China and the United States: Cross-cultural insights into the nature and course of preschool mathematical development," in *Handbook of Mathematical Cognition*, ed. J. I. D. Campbell (New York, NY: Psychology Press), 163–178.

Mousley, K., and Kurz, C. (2015). Pre-college deaf students' understanding of fractional concepts: What we know and what we do not know. *J. Sci. Educ. Stud. Disabil.* 18, 44–65.

Noël, M.-P., Seron, X., and Trovarelli, F. (2004). Working memory as a predictor of addition skills and addition strategies in children. *Curr. Psychol. Cogn.* 22, 3–25.

Nunes, T., Bryant, P., Burman, D., Bell, D., Evans, D., and Hallett, D. (2009). Deaf children's informal knowledge of multiplicative reasoning. *J. Deaf Stud. Deaf Educ.* 14, 260–277. doi: 10.1093/deafed/enn040

Nunes, T., and Moreno, C. (2002). An intervention program for promoting deaf pupils' achievement in mathematics. *J. Deaf Stud. Deaf Educ.* 7, 120–133. doi: 10.1093/deafed/7.2.120

Núñez-Peña, M. I., Gracia-Bafalluy, M., and Tubau, E. (2011). Individual differences in arithmetic skill reflected in event-related brain potentials. *Int. J. Psychophysiol.* 80, 143–149. doi: 10.1016/j.ijpsycho.2011.02.017

Oberauer, K., and Lange, E. B. (2008). Interference in verbal working memory: Distinguishing similarity-based confusion, feature overwriting, and feature migration. *J. Mem. Lang.* 58, 730–745. doi: 10.1016/j.jml.2007.09.006

Pagliaro, C. M. (2010). "Mathematics instruction and learning of deaf and hard-of-hearing students: What do we know? Where do we go," in *The Oxford Handbook of Deaf Studies, Language, and Education*, Vol. 2, eds M. Marschark and P. E. Spencer (Cary, NC: Oxford University Press), 156–171.

Pagliaro, C. M., and Ansell, E. (2012). Deaf and hard of hearing students' problem-solving strategies with signed arithmetic story problems. *Am. Ann. Deaf* 156, 438–458. doi: 10.1353/aad.2012.1600

Pagliaro, C. M., and Kritzer, K. L. (2013). The math gap: A description of the mathematics performance of preschool-aged deaf/hard-of-hearing children. *J. Deaf Stud. Deaf. Educ.* 18, 139–160. doi: 10.1093/deafed/ens070

Paivio, A., and Csapo, K. (1971). Short-term sequential memory for pictures and words. *Psychon. Sci.* 24, 50–51. doi: 10.3758/BF03337887

Pasolunghi, M. C., Cornoldi, C., and De Liberto, S. (1999). Working memory and intrusions of irrelevant information in a group of specific poor problem solvers. *Mem. Cogn.* 27, 779–790. doi: 10.3758/BF03198531

Passolunghi, M. C., and Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. *J. Exp. Psychol.* 88, 348–367. doi: 10.1016/j.jecp.2004.04.002

Prado, J., Mutreja, R., and Booth, J. R. (2014). Developmental dissociation in the neural responses to simple multiplication and subtraction problems. *Dev. Sci.* 17, 537–552. doi: 10.1111/desc.12140

Prado, J., Mutreja, R., Zhang, H., Mehta, R., Desroches, A. S., Minas, J. E., et al. (2011). Distinct representations of subtraction and multiplication in the neural systems for numerosity and language. *Hum. Brain Mapp.* 32, 1932–1947. doi: 10.1002/hbm.21159

Quandt, L. C., and Kubicek, E. (2018). Sensorimotor characteristics of sign translations modulate EEG when deaf signers read English. *Brain Lang.* 187, 9–17. doi: 10.1016/j.bandl.2018.10.001

Rasmussen, C., Ho, E., Nicoladis, E., Leung, J., and Bisanz, J. (2006). Is the Chinese number-naming system transparent? Evidence from Chinese-English bilingual children. *Can. J. Exp. Psychol.* 60, 60–67. doi: 10.1037/cjep2006007

Ribner, A. D., Peretz-Lange, R., Berkowitz, T., and Shusterman, A. (2022). The role of language in executive function development: Evidence from oral deaf preschoolers. *PsyArXiv* [Preprint], doi: 10.31234/osf.io/8ahm9

Robinson, K. M. (2001). The validity of verbal reports in children's subtraction. *J. Educ. Psychol.* 93, 211–222.

Rodriguez-Santos, J. M., Calleja, M., Garcia-Orza, J., Iza, M., and Damas, J. (2014). Quantity processing in deaf and hard of hearing children: Evidence from symbolic and nonsymbolic comparison tasks. *Am. Ann. Deaf.* 159, 34–44. doi: 10.1353/aad.2014.0015

Rudner, M., Orfanidou, E., Cardin, V., Capek, C. M., Woll, B., and Rönnberg, J. (2012). "Reading ability in adult deaf native signers is positively associated with their ability to judge the grammatically of their native sign language," in *Neurobiology of Language Conference*, San Sebastian.

Salvaggio, S., Masson, N., Zénon, A., and Andres, M. (2022). The predictive role of eye movements in mental arithmetic. *Exp. Brain Res.* 240, 1331–1340. doi: 10.1007/s00221-022-06329-3

Sandler, W., and Lillo-Martin, D. C. (2006). *Sign Language and Linguistic Universals*. Cambridge, NY: Cambridge University Press.

Sandrini, M., Miozzo, A., Cotelli, M., and Cappa, S. F. (2003). The residual calculation abilities of a patient with severe aphasia: Evidence for a selective deficit of subtraction procedures. *Cortex* 39, 85–96. doi: 10.1016/s0010-9452(08)70076-5

Santos, S., and Cordes, S. (2022). Math abilities in deaf and hard of hearing children: The role of language in developing number concepts. *Psychol. Rev.* 129, 199–211. doi: 10.1037/rev0000303

Serrano Pau, C. (1995). The deaf child and solving problems of arithmetic. The importance of comprehensive reading. *Am. Ann. Deaf* 140, 287–290. doi: 10.1353/aad.2012.0599

Seyler, D. J., Kirk, E. P., and Ashcraft, M. H. (2003). Elementary subtraction. *J. Exp. Psychol.* 29, 1339–1352.

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *J. Exp. Psychol. Gen.* 117, 258–275. doi: 10.1037//0096-3445.117.3.258

Simmons, F. R., and Singleton, C. (2008). Do weak phonological representations impact on arithmetic development? A review of research into arithmetic and dyslexia. *Dyslexia* 14, 77–94. doi: 10.1002/dys.341

Stokoe, W., Casterline, D., and Croneberg, C. (1965). *A Dictionary of ASL on Linguistic Principles*. Washington, DC: Gallaudet College Press.

Temple, C. M., and Sherwood, S. (2002). Representation and retrieval of arithmetical facts: Developmental difficulties. *Q. J. Exp. Psychol. Sect. A* 55, 733–752. doi: 10.1080/02724980143000550

Thevenot, C., and Barrouillet, P. (2006). Encoding numbers: Behavioral evidence for processing-specific representations. *Mem. Cogn.* 34, 938–948. doi: 10.3758/bf03193439

Thevenot, C., Barrouillet, P., and Fayol, M. (2001). Algorithmic solution of arithmetic problems and operands-answer associations in long-term memory. *Q. J. Exp. Psychol. A* 54, 599–611. doi: 10.1080/713755966

Thevenot, C., Fanget, M., and Fayol, M. (2007). Retrieval or nonretrieval strategies in mental arithmetic? An operand recognition paradigm. *Mem. Cogn.* 35, 1344–1352. doi: 10.3758/BF03193606

Titus, J. C. (1995). The concept of fractional number among deaf and hard of hearing students. *Am. Ann. Deaf* 140, 255–262. doi: 10.1353/aad.2012.0582

van Harskamp, N. J., and Cipolotti, L. (2001). Selective impairments for addition, subtraction and multiplication. implications for the organisation of arithmetical facts. *Cortex* 37, 363–388. doi: 10.1016/s0010-9452(08)70579-3

Van Rinsveld, A., Brunner, M., Landerl, K., Schiltz, C., and Ugen, S. (2015). The relation between language and arithmetic in bilinguals: Insights from different stages of language acquisition. *Front. Psychol.* 6:265. doi: 10.3389/fpsyg.2015.00265

Van Rinsveld, A., Dricot, L., Guillaume, M., Rossion, B., and Schiltz, C. (2017). Mental arithmetic in the bilingual brain: Language matters. *Neuropsychologia* 101, 17–29. doi: 10.1016/j.neuropsychologia.2017.05.009

Van Rinsveld, A., Schiltz, C., Landerl, K., Brunner, M., and Ugen, S. (2016). Speaking two languages with different number naming systems: What implications for magnitude judgments in bilinguals at different stages of language acquisition? *Cogn. Process.* 17, 225–241. doi: 10.1007/s10339-016-0762-9

Verguts, T., and Fias, W. (2005). Interacting neighbors: A connectionist model of retrieval in single-digit multiplication. *Mem. Cogn.* 33, 1–16. doi: 10.3758/BF03195293

Wang, J., Lin, E., Tanase, M., and Sas, M. (2008). Revisiting the influence of numerical language characteristics on mathematics achievement: Comparison among China, Romania, and U.S. *Int. Electr. J. Math. Educ.* 3, 24–46. doi: 10.29333/iejme/216

Watkins, M. J., LeCompte, D. C., Elliott, M. N., and Fish, S. B. (1992). Short-term memory for the timing of auditory and visual signals. *J. Exp. Psychol. Learn. Mem. Cogn.* 18, 931–937. doi: 10.1037//0278-7393.18.5.931

Watkins, O. C., and Watkins, M. J. (1980). The modality effect and echoic persistence. *J. Exp. Psychol. Gen.* 109, 251–278. doi: 10.1037//0096-3445.109.3.251

Wickelgren, W. A. (1979). *Cognitive Psychology*. Englewood Cliffs, NJ: Prentice-Hall.

Wilson, M. (2001). The case for sensorimotor coding in working memory. *Psychon. Bull. Rev.* 8, 44–57. doi: 10.3758/bf03196138

Wilson, M., and Emmorey, K. (1997). A visuospatial "phonological loop" in working memory: Evidence from American sign language. *Mem. Cogn.* 25, 313–320. doi: 10.3758/BF03211287

Wollman, D. C. (1965). The attainments in English and arithmetic of secondary school pupils with impaired hearing. *Teach. Deaf* 159, 121–129.

Wood, D., Wood, H., Griffith, A., and Howarth, I. (1986). *Teaching and Talking With Deaf Children*. New York, NY: Wiley.

Zhou, X., Chen, C., Zang, Y., Dong, Q., Chen, C., Qiao, S., et al. (2007). Dissociated brain organization for single-digit addition and multiplication. *Neuroimage* 35, 871–880.

frontiers | Frontiers in Human Neuroscience

TYPE Original Research
PUBLISHED 23 December 2022
DOI 10.3389/fnhum.2022.1062168

OPEN ACCESS

EDITED BY
Elena Nava,
University of Milano-Bicocca, Italy

REVIEWED BY
Erkin Asutay,
Linköping University, Sweden
Poppy Watson,
University of New South Wales,
Australia

*CORRESPONDENCE
Jessica Emily Antono
✉ jessica.e.antono@gmail.com
Arezoo Pooresmaeili
✉ arezoo.pooresmaeili@gmail.com

SPECIALTY SECTION
This article was submitted to
Sensory Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 05 October 2022
ACCEPTED 06 December 2022
PUBLISHED 23 December 2022

CITATION
Antono JE, Vakhrushev R and
Pooresmaeili A (2022) Value-driven
modulation of visual perception by
visual and auditory reward cues:
The role of performance-contingent
delivery of reward.
Front. Hum. Neurosci. 16:1062168.
doi: 10.3389/fnhum.2022.1062168

Value-driven modulation of visual perception by visual and auditory reward cues: The role of performance-contingent delivery of reward

Jessica Emily Antono*, Roman Vakhrushev and Arezoo Pooresmaeili*

Perception and Cognition Lab, European Neuroscience Institute Göttingen–A Joint Initiative of the University Medical Center Göttingen and the Max-Planck-Society, Göttingen, Germany

Perception is modulated by reward value, an effect elicited not only by stimuli that are predictive of performance-contingent delivery of reward (PC) but also by stimuli that were previously rewarded (PR). PC and PR cues may engage different mechanisms relying on goal-driven versus stimulus-driven prioritization of high value stimuli, respectively. However, these two modes of reward modulation have not been systematically compared against each other. This study employed a behavioral paradigm where participants' visual orientation discrimination was tested in the presence of task-irrelevant visual or auditory reward cues. In the first phase (PC), correct performance led to a high or low monetary reward dependent on the identity of visual or auditory cues. In the subsequent phase (PR), visual or auditory cues were not followed by reward delivery anymore. We hypothesized that PC cues have a stronger modulatory effect on visual discrimination and pupil responses compared to PR cues. We found an overall larger task-evoked pupil dilation in PC compared to PR phase. Whereas PC and PR cues both increased the accuracy of visual discrimination, value-driven acceleration of reaction times (RTs) and pupillary responses only occurred for PC cues. The modulation of pupil size by high reward PC cues was strongly correlated with the modulation of a combined measure of speed and accuracy. These results indicate that although value-driven modulation of perception can occur even when reward delivery is halted, stronger goal-driven control elicited by PC reward cues additionally results in a more efficient balance between accuracy and speed of perceptual choices.

KEYWORDS

visual perception, reward, pupil response, sensory modality, psychophysics

Introduction

Stimuli associated with rewards have a strong influence on our behavior as they trigger the expectation of desirable outcomes, thereby driving agents to optimize their goal-directed actions (Schultz, 2015) and value-based choices (Delgado, 2007; Wallis, 2007; Schultz, 2015). Accordingly, brain areas underlying action planning and value-based decisions are strongly modulated by rewards. Moreover, reward effects even extend to the earliest stages of information processing in the brain as reward associations of stimuli influence their representation in the primary sensory areas (Shuler and Bear, 2006; Serences, 2008). Understanding the underlying mechanisms of value-driven modulation of perception is important since it allows a better understanding of how experience-related and contextual factors in general influence sensory perception (Pessoa and Engelmann, 2010; Seriès and Seitz, 2013).

Reward effects on perception are typically investigated using paradigms where correct detection or discrimination in a perceptual task (Engelmann and Pessoa, 2007) or efficient orienting responses in a motor task (Milstein and Dorris, 2007) lead to higher magnitude or probability of rewards. In such scenarios, prioritization of reward cues, through engaging mechanisms such as selective attention or preparation of oculomotor responses, aligns with the goal-driven mechanisms that help agents to maximize their obtained rewards (Chelazzi et al., 2013; Failing and Theeuwes, 2018). Using such tasks, value-driven modulations have been observed at the early stages of sensory processing in the brain. For instance, Weil et al. (2010) provided evidence that rewarding feedbacks improved behavioral performance in a visual discrimination task and also increased the activity in the human primary visual cortex during the discrimination phase following a reward feedback. Another study by Pleger et al. (2008) also demonstrated that reward facilitated somatosensory judgments. There, high reward cues improved tactile performance and enhanced the hemodynamic response in the primary somatosensory cortex, indicating that reward signals can influence early sensory areas when a decision is based on the sensory features of stimuli. Thus, reward signals, during the delivery of reward or during the presentation of reward-predicting cues, can be propagated not only within the classical reward-related regions, but also to sensory areas, especially when the reward delivery is contingent on the accuracy of sensory judgments [i.e., performance-contingent (PC)]. One criticism to these designs is that value-driven effects cannot be distinguished from attentional (Maunsell, 2004) or cognitive control mechanisms (Botvinick and Braver, 2015) that are involved in processing of the task-relevant feature of a task. Accordingly, such paradigms do not allow a differentiation between value-driven effects due to voluntary, goal-driven mechanisms from effects due to stimulus-driven and involuntary mechanisms.

Another line of research has shown that value-driven modulation of perception also occurs when reward cues are not the relevant feature of the task or when reward delivery and hence the motivation to strategically optimize performance has been removed. For instance, the delivery of reward in response to a saccadic target in some trials can affect the oculomotor performance in subsequent unrewarded trials when a non-target stimulus contains a similar feature as the rewarded target in the past (Hickey and van Zoest, 2012). It has also been shown that reward effects outlast the delivery of reward so that previously rewarded (PR) features automatically affect participants' performance (Yantis et al., 2012; De Tommaso et al., 2017). The latter experiments typically employ a two-phase paradigm (De Tommaso and Turatto, 2021), where in the first training or conditioning phase participants learn the association of stimulus features with certain amount or probability of reward, and in the subsequent test phase PR cues are presented without the actual delivery of reward (i.e., during extinction). Although during the test phase reward associated cues are not reinforced anymore, it has consistently been shown that they can still involuntarily capture participants' attention, a phenomenon called value-driven attentional capture (VDAC) (Anderson et al., 2011), and thereby influence perceptual judgments across a variety of tasks (Anderson et al., 2011; Yantis et al., 2012; Camara et al., 2013; Failing and Theeuwes, 2015; Bucker and Theeuwes, 2017; Tankelevitch et al., 2020). The typical finding of these studies is that when PR stimuli are the same as the target of a task they facilitate performance (accuracy or RT) but importantly when they are irrelevant to the task or assigned to distractors, they can impair performance (Anderson et al., 2014; Asutay and Västfjäll, 2016; Gong et al., 2017; Bucker and Theeuwes, 2018; Qin et al., 2020; Watson et al., 2020), a so-called value-driven distraction (Rusz et al., 2020). Such effects likely arise as a result of the enhanced representation of distractors in visual cortex (Itthipuripat et al., 2019), which limit the processing resources that are available to the target.

Interestingly, it is not always the case that task-irrelevant reward cues capture attention away from the target and suppress performance. For instance, Pooresmaeili et al. (2014) utilized one sensory modality (audition) to signal the reward value while keeping the target of the task in another modality (vision). Using this design, it was shown that task-irrelevant auditory cues that were previously associated with high reward enhanced the visual sensitivity compared to low reward cues. A follow-up study (Vakhrushev et al., 2021) used a similar design and compared task-irrelevant reward cues from the same (vision) or different (audition) sensory modality in terms of their effect on perceptual decisions made about a visual target. In this study, it was found that PR auditory and visual cues had distinct effects on behavioral and electrophysiological correlates of visual perception, suggesting that reward-driven modulations may have dependencies on the sensory modality of task-irrelevant stimuli.

Overall, across different paradigms employed to investigate the effects of reward on sensory perception, PC rewards have been often found to be associated with the facilitation of sensory processing, whereas divergent effects were observed for cues previously associated with rewards based on whether the target or the task-irrelevant distractors contained a rewarded feature. Another factor that also seems to weigh in is where the reward information was signaled from, with different effects for rewards cued intra-modally or cross-modally. However, a systematic investigation of these factors where the same perceptual judgment is tested under different modes of reward delivery and cuing has been missing. Therefore, in the current study, we designed a paradigm that tested the effect of three factors on visual perception: reward magnitude, sensory modalities of reward cues, and the contingency of reward delivery on task performance. Specifically, a similar design as two previous studies from our lab (Pooresmaeili et al., 2014; Vakhrushev et al., 2021) was used where auditory or visual cues were first associated with either high or low monetary reward during a training phase (referred to as conditioning). During the test phase, auditory and visual cues were presented at the same time as the target of a visual discrimination task but did not carry any information about the task at hand (i.e., orientation discrimination). Importantly, participants either obtained rewards upon correct responses or did not receive any reward feedback in any condition. In the first case, participants' rewards depended on the identity of auditory or visual stimuli and these cues were *PC* predictors of rewards, whereas in the second case auditory and visual stimuli were *previously associated with rewards* (PR) and did not predict the delivery of reward anymore. We hypothesized the two modes of reward cuing are linked to distinct processes: goal-driven (voluntary) and stimulus-driven (involuntary) attention. In result, when the cues were PC, the voluntary control would dominate and therefore the cues would benefit performance. However, when the cues were associated with rewards in the past and did not lead to reward feedbacks during the test phase, they would only involve the involuntary capture of attention and lead to weaker reward-driven modulations, which may differ between the intra- and cross-modal rewards. Pupil responses can be used as a sensitive readout of changes in the motivational state due to salient events (Chiew and Braver, 2013; Schneider et al., 2018; Pietrock et al., 2019), even when such events are not consciously detected (Bijleveld et al., 2009). Pupil responses have also been recently linked to the level of cognitive effort exerted in a task (van der Wel and van Steenbergen, 2018). We therefore hypothesized that PC reward cues are associated with higher goal-directed cognitive effort in prospect of higher rewards, hence producing a stronger value-driven modulation of pupillary responses compared to cues that were previously associated with rewards.

Our results demonstrate that reward associated cues enhance the accuracy of visual discrimination irrespective of the sensory modality and whether the reward delivery was continued (PC) or halted (PR). Additionally, PC reward cues energized behavior, as indexed by reaction times (RTs) and pupil responses, an effect that was absent in PR cues.

Materials and methods

Participants

In total, 43 subjects participated in the experiment to fulfill a target sample size of $N = 36$ based on a previous study (Vakhrushev et al., 2021). They were invited *via* an online recruiting system.[1] All participants were naïve to the hypothesis of the project, had no history of neurophysiological or psychiatric disorders according to a self-report, had normal or corrected-to-normal vision, and performed the key presses during the task with their dominant hands (five left handed). Eight participants were removed from the final sample, as due to technical problems the experiment had to be terminated before the complete dataset was collected ($N = 4$), the psychometric method used to estimate the orientation discrimination thresholds did not converge on a reliable value ($N = 2$, based on our previous work the QUEST method needed to converge on a stimulus orientation $< 2°$ and performance during the baseline phase needed to be $<90\%$), the participant did not learn the reward associations ($N = 1$) or had a strong bias for one of the colors or sounds prior to learning the reward associations ($N = 1$, estimated as a bias toward high reward colors or sounds > 2.5 SD of the group mean). Thus, the final sample comprised data from 35 participants (18 female; age: 18–45, 27 ± 5 SD years).

Participants were informed that after the experiment they would obtain a reward comprising a fixed hourly rate (~8 Euros per hours) plus an added bonus that depended on their performance. To calculate the total reward, the fixed hourly rate was added to the money participants obtained during the experiment and a fraction of the total amount (4%) was handed over to the participants in cash.

Before the experiment started and after all procedures were explained, participants gave their oral and written consent. The study was approved by the local ethics committee of the "Universitätsmedizin Göttingen" (UMG), under the proposal number 15/7/15.

Stimulus presentation and apparatus

The behavioral paradigms used during the reward associative learning (conditioning) and test phase were

---

[1] http://www.probanden.eni-g.de/orsee/public/

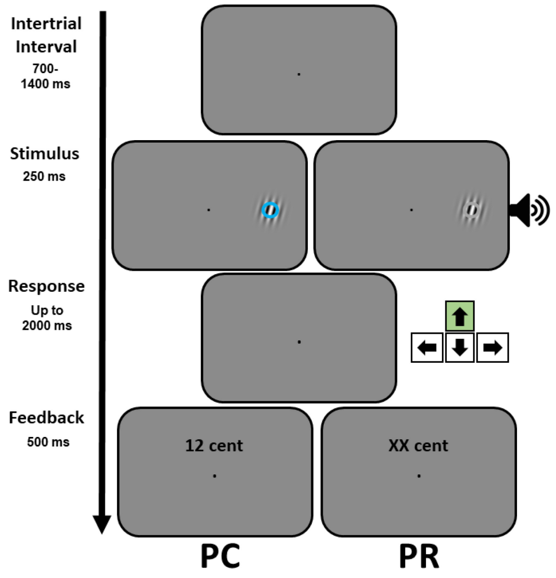


FIGURE 1
Behavioral paradigm employed during the test phase. An example trial of the visual discrimination task, illustrating the Gabor target and the task-irrelevant visual **(left)** or auditory **(right)** stimuli, is shown. Participants reported the orientation of the Gabor target by pressing either the up or down arrow keys (the correct response for the example trial is illustrated symbolically by the arrow in the green box). Prior to the test phase, participants learned to associate different visual (blue or orange circles) or auditory (high or low pitch tones) stimuli, counter-balanced across participants, with different reward magnitudes during a conditioning phase (see **Supplementary Figure 1**). The test phase comprised two parts with different reward contingencies (PC and PR). In case of a correct response, during the performance-contingent reward (PC) phase, the monetary reward associated with a specific stimulus was displayed (for instance 12 cent). In a subsequent phase, previously reward-associated (PR) stimuli were not predictive of reward delivery, but to keep the layout of the feedback display similar across the two phases the letters XX were shown for all conditions.

identical to a previous study (Vakhrushev et al., 2021). The paradigm employed during the conditioning was a spatial localization task (see **Supplementary Figure 1** and the Section "Experimental procedure") where participants reported the side (left or right) from which visual or auditory stimuli were presented. During the test phase, a visual orientation discrimination task was used in which the tilt direction of a Gabor patch (a Gaussian-windowed sinusoidal grating with SD = 0.33°, a spatial frequency of 3 cycles per degree, subtending 2° diameter, displayed at 9° eccentricity to the left or right side of the fixation point) had to be reported (**Figure 1**). The tilt orientation of the Gabor patch was set to each participant's perceptual threshold estimated after the initial training. To determine this threshold, we employed a QUEST algorithm (Watson and Pelli, 1983) to estimate the Gabor tilt orientation for which participants' performance was at 70%. In each trial, a task-irrelevant semi-transparent ring (alpha 50%, 0.44° in diameter) was superimposed on the Gabor patch. The color of the rings (orange or blue for visual conditions, or gray for auditory and neutral conditions) was adjusted individually for each participant in such a way that they were perceptually isoluminant. Perceptual thresholds for the visual discrimination task were determined when Gabors were superimposed with a gray circle. For auditory cues, two pure tones with different frequencies (350 or 1,050 Hz) were presented at 70 dB simultaneously with the Gabor patch and at the same side.

The timing of events was identical across the experiment (see **Figure 1** and **Supplementary Figure 1**). As soon as participants fixated (within 1° of the fixation point) a trial started. After an additional fixation period of 700–1400 ms, a target stimulus appeared (either a colored circle or a tone during conditioning or a Gabor patch together with a colored circle or a tone during the test phase). The target stimulus disappeared after 250 ms and participants had to indicate its side (conditioning) or the orientation of the Gabor patch (during the test phase) within 2,000 ms from the onset of the target. Finally, a feedback display was presented for 500 ms. The feedback display contained the reward magnitude that participants received (in numbers) during conditioning and PC phase (see the Section "Experimental procedure"). To keep the visual layout of the feedback display similar across PC and PR phases, in the latter phase "xx cent" was shown for all conditions.

Throughout the experiment, visual stimuli were displayed on a calibrated ViewPixx monitor (refresh rate = 120 Hz, resolution 1,080 × 1,920 pixels, and placed at a viewing distance of 60 cm). The auditory tones were delivered through an over-ear headphone (HAD 280 audiometry headphones, Sennheiser).

Experimental procedure

The experiment consisted of a practice session (32 trials) for the orientation discrimination task and three phases. In the first phase, referred to as the *baseline phase* (160 trials), participants were required to report the tilt direction of a Gabor patch relative to the horizontal meridian by pressing a keyboard button (either the down or up arrow keyboard button for clockwise and counter-clockwise directions, respectively; see **Figure 1**). They were additionally instructed to ignore the simultaneously presented visual or auditory cues that accompanied the Gabor. Afterward, participants completed a *conditioning* task to learn the reward associations of auditory and visual cues (see **Supplementary Figure 1**). In this task, participants decided whether a colored circle or an auditory tone was perceived to be on the left or right side by pressing the corresponding arrow key buttons. Upon correct response, participants saw the magnitude of the reward that was paired with a certain cue and thereby learned whether a visual or auditory stimulus was associated with high (mean = 25 Cents) or low (mean = 2 Cents, drawn from a Poisson distribution) monetary reward. In the third phase, referred as the *test phase*, participants performed the same orientation discrimination task as in the baseline phase, but in the presence of task-irrelevant visual or auditory cues that had been associated with different amounts of reward during conditioning. As the main task was a visual discrimination task, task-irrelevant visual and auditory stimuli will be referred to as intra- and cross-modal, respectively. Additionally, the test phase was split into two parts: in the first part (320 trials, the phase with *performance-contingent reward cues*, *PC*), upon correct response, similar reward feedbacks as in the conditioning phase were presented, i.e., reward depended on the identity of cues and was either high or low. In the second part (320 trials, referred to as the phase with *previously associated reward cues*, *PR*), the delivery of rewards was halted. Here, participants were instructed similarly to the PC phase with the exception that they were informed about a different feedback display shown after each trial. Specifically, they were told that in the PR phase the differential reward deliveries would be halted and instead after each trial they would see a feedback in the form of "xx cent" indicating a constant amount of reward that would be added to their total earning in case they responded correctly.

In order to determine whether participants learned the reward-cue association, they were asked to indicate which cue from each modality presented to them sequentially had been associated with more money. This question was completed in multiple parts following the conditioning, PC, and PR phases. Additionally, we also repeated the question in the questionnaire after the experiment was completed. If a participant did not provide any correct response across all experimental phases (*conditioning*, PC and PR), then the participant was removed from further analysis ($N = 1$).

Pupillometry

An EyeLink 1000 Plus system with a desktop mount (SR Research) was used to track the right eye. The EyeLink camera was controlled by the corresponding toolbox in MATLAB (Cornelissen et al., 2002). Before each block, the eye tracking system was calibrated using a nine-point standard EyeLink calibration procedure.

Pupil responses were acquired at a sampling frequency of 1,000 Hz. The pupil data of each trial was extracted from 100 ms prior to the target onset until the end of the trial (i.e., the end of the feedback display). Trials in which more than 50% of data was lost were removed from further analysis. For the missing data due to blinks, a linear interpolation was applied, where the missing data was interpolated based on the samples within a window of 10 ms before and after the blink. The data was then low-pass filtered (fourth order Butterworth with a cut-off frequency of 2 Hz), normalized to z-score (across all samples recorded for each participant) and subsequently corrected for baseline (i.e., 100 ms). For the statistical analysis, the average stimulus-evoked response in a window from the target onset until the end of each trial (the end of the feedback display as shown in **Figure 1**) was examined. Note that a trial's timing depended on how fast the participant responded. Therefore, to examine the relation between the pupil size and the behavioral measures, pupil responses were estimated from the data of the first 500 ms interval after the target onset. This was done to ensure that for all participants and all experimental conditions the same number of pupil samples were considered.

Data analysis

The data obtained from all parts of the experiment was analyzed using custom-written scripts in MATLAB (version R2015a). We analyzed accuracies, reaction times (RT: median reaction time across correct and incorrect trials), inverse efficiency scores (IE) (median RT of correct trials divided by the accuracy) d-prime (d') and pupil size. We removed trials in which any of the following conditions were met: lack of stable fixation during the presentation of the target (i.e., the distance of eye gaze from the fixation point exceeded 0.9°), no response, RTs exceeding the 2.5 SD of each phase, or loss of more than 50% of pupil data. This resulted in 2.98% (±1.20 SD), 2.62% (±2.25 SD), 3.01% (±1.04 SD), and 3.64% (±2.97 SD) trials removed from baseline, conditioning, PC and PR phases, respectively. For each response variable, we calculated the average across all trials of each condition per subject during the baseline and test phases separately. D-prime was measured based on the probability of hits and false-alarms, as $d' = Z(PHit)–Z(PFA)$, where one of the tilt directions was arbitrarily treated as "target-present" as in formal Signal Detection Theory analysis of discrimination tasks (Macmillan and Creelman, 1991). Extreme values of PHit or PFA were slightly up- or down-adjusted (i.e., a probability equal to 0 or 1 was adjusted by adding or subtracting $\frac{1}{2 \times N}$, where N is the number of trials, respectively). Afterward, the difference in response variables (accuracies, reaction times, d' and pupil size) between baseline and test phase was entered to a 2 × 2 × 2 repeated measures ANOVA, with the reward contingency (performance-contingent: PC and previously associated: PR), reward magnitude (high and low), and sensory modality (visual or auditory, i.e., intra- and cross-modal, respectively) as within-subjects factors. Significant effects in RM ANOVA were followed up by *post-hoc* tests (*multcompare* in MATLAB with *Bonferroni* correction). To test whether the value-driven modulation of pupil size is predictive of the modulation of the behavioral measures a robust regression method (*robustfit* with default settings in MATLAB) was employed.

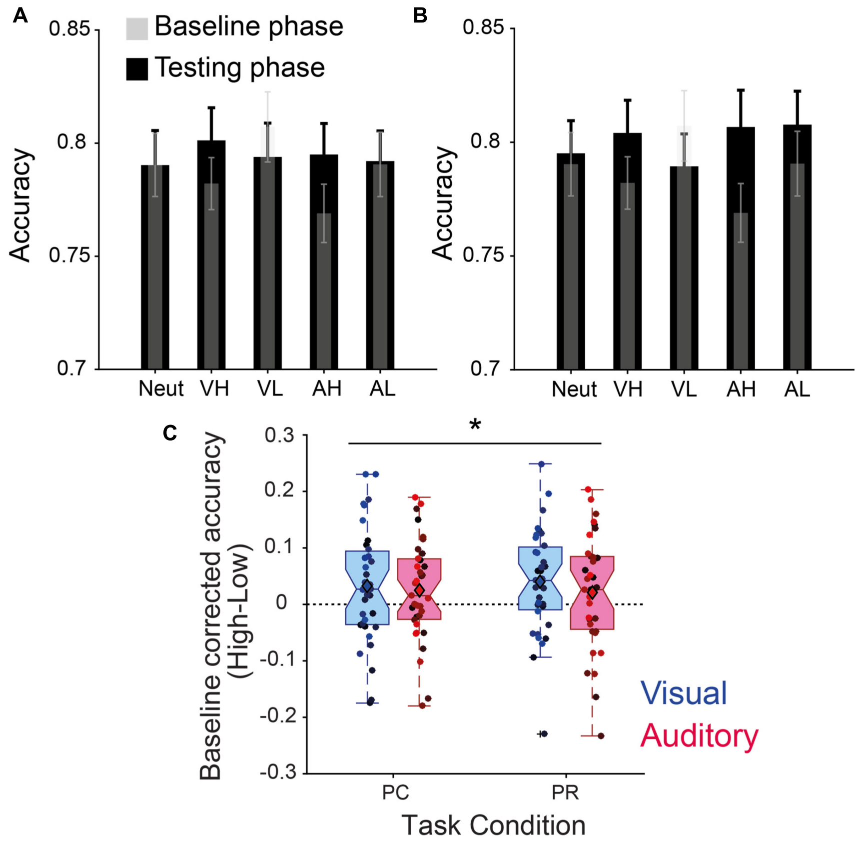


FIGURE 2

Value-driven modulation of discrimination accuracy. **(A)** Accuracies of the baseline and performance-contingent reward (PC) phase. **(B)** Same as panel **(A)** for the previously associated rewards (PR) phase. The transparent gray shades represent the baseline performance before learning the reward associations, overlaid on the test phase performance in black for each condition (neut, neutral; VH, visual high-; VL, visual low-; AH, auditory high-; and AL, auditory low-reward). **(C)** Baseline -corrected reward effect (high−low) for intra-modal (visual) and cross-modal (auditory) reward cues during the two phases. Error bars in panels **(A,B)** represent s.e.m., circles with different color shades in panel **(C)** correspond to the data of individual participants, and * stands for the main effect of reward at $p < 0.05$.

Results

The main objective of this study was to examine whether visual discrimination is influenced by co-occurring visual and auditory stimuli which did not carry any information about the dimension over which the discrimination was performed (i.e., the orientation of a Gabor stimulus, see **Figure 1**) but were either predictive of the reward delivery upon correct performance (i.e., performance-contingent: PC phase) or were previously associated with the reward delivery (i.e., previously rewarded: PR phase). Participants first learned the reward associations of visual and auditory stimuli during a conditioning phase by performing a localization task (see the **Supplementary Text** and **Supplementary Figure 1**). We found a weak effect of reward on the behavioral performance and pupil responses (see the **Supplementary Text** and **Supplementary Figure 2**) during the conditioning phase. Nevertheless, the conditioning task was successful in establishing the associations between stimuli and rewards, as according to the debriefings performed after this phase, all participants had learned the reward associations of tones and colors correctly. Therefore, we next examined the behavioral and pupillometric responses during the visual discrimination task, testing whether the learned reward associations affected the visual perception during the PC and PR phases compared to the baseline (i.e., done prior to the conditioning).

Effect of performance-contingent and previously associated reward cues on the accuracy of visual discrimination

Overall, during the initial baseline phase where the cues were not associated with any reward magnitude, participants performed on average across all conditions with 78.78% accuracy (±0.94 s.e.m) (**Figures 2A, B**), while in the PC phase, mean accuracy increased to 79.44% (±1.23 s.e.m) and in the last phase with PR cues increased to 80.06% (±1.32 s.e.m). This indicated that with time, participants became more proficient in the task. However, the improvement of accuracy across time (Baseline, PC and PR) did not reach statistical significance [$F(2,34) = 1.04$, $p = 0.35$, $\eta_p^2 = 0.03$].

In the test phase, a repeated measures 2 × 2 × 2 ANOVA conducted on the baseline corrected accuracy rates showed a significant main effect of reward magnitude across PC and PR phases (**Figure 2C**): $F(1,34) = 7.37$, $p = 0.01$, $\eta_p^2 = 0.18$. All other main and interaction effects were non-significant (all $ps > 0.1$). *Post-hoc* tests revealed a significant increase in accuracies by high- compared to low-reward visual cues in PR ($p = 0.016$, Cohen's d = 0.430), a trend in PC ($p = 0.068$, Cohen's d = 0.319) and non-significant effects in auditory conditions (PC: $p = 0.108$, Cohen's d = 0.279; and PR: $p = 0.235$, Cohen's d = 0.204). We obtained similar results when d-prime (d′) scores instead of accuracies were used [$F(1,34) = 6.75$, $p = 0.01$, $\eta_p^2 = 0.17$], indicating that the improvement in participants' performance was not driven by an enhanced false-alarm rate.

The main effect of reward is in line with our hypothesis predicting that high-reward cues improve the perceptual discriminability. Contrary to our predictions, we did not find a significant interaction effect with reward contingency or sensory modality, although the effect sizes were larger for intra-modal (visual) cues.

Effect of performance-contingent and previously associated reward cues on the speed of visual discrimination

The analysis of RTs across all conditions demonstrated that participants became overall faster as they proceeded through the experiment (**Figures 3A, B**), an effect that reached statistical significance when tested with an ANOVA with phase (Baseline, PC and PR) as the independent factor [$F(2,34) = 21.39$, $p < 10^{-7}$, $\eta_p^2 = 0.39$]. Participants' RTs in both PC ($M = 770.83$ ms, s.e.m = 18.24 ms) and PR phases ($M = 782.41$ ms, s.e.m = 18.93 ms) were significantly faster than the baseline phase ($M = 843.01$ ms, s.e.m = 21.33 ms, both $ps < 10^{-4}$).

A repeated measures 2 × 2 × 2 ANOVA on the baseline corrected RTs revealed a significant interaction between reward magnitude and task contingency [$F(1,34) = 4.61$, $p = 0.039$, $\eta_p^2 = 0.12$, **Figure 3C**]. This effect demonstrates that when cues associated with higher value were predictive of the reward delivery, participants reacted faster than when reward delivery was halted. Specifically, *post-hoc* tests revealed that this effect was more pronounced for PC, high-reward visual cues ($p = 0.048$, Cohen's d = 0.33) than other conditions (visual/PR: $p = 0.47$, Cohen's d = 0.123; auditory/PR: $p = 0.30$, Cohen's d = 0.178; auditory/PC: $p = 0.80$, Cohen's d = 0.043). Although mostly driven by the visual cues, this finding is in line with our hypothesis predicting that PC rewards have a stronger influence on the speed of perceptual decisions.

Effect of performance-contingent and previously associated reward cues on pupil responses

We next examined the pupil responses using a 2 × 2 × 2 repeated measure ANOVA with three factors: reward magnitude (high and low), sensory modality (auditory and visual), and reward contingency (performance-contingent: PC and previously associated: PR). Pupil responses were the baseline corrected average pupil size (z-score) extracted from the target onset until the trial end (**Figure 4**). Across all visual and auditory conditions, task-evoked pupil responses were significantly higher in PC compared to PR phase [$F(1,34) = 61.32$, $p < 10^{-8}$, $\eta_p^2 = 0.643$]. Additionally, a significant interaction effect was observed between the reward magnitude and contingency [$F(1,34) = 7.17$, $p = 0.011$, $\eta_p^2 = 0.174$], as higher rewards increased the pupil size compared to lower rewards only in PC ($p = 0.04$, Cohen's d = 0.354) but not in PR phase ($p = 0.94$, Cohen's d = 0.014). A weaker interaction effect [$F(1,34) = 4.80$, $p = 0.035$, $\eta_p^2 = 0.124$] was also observed between the sensory modality and reward contingency, corresponding to larger pupil responses evoked by cross-modal (auditory) compared to intra-modal (visual) stimuli in PC phase and an opposite effect in PR phase. The effect of sensory modality in each phase did not reach significance (PC: auditory-visual = 0.02 ± 0.02 s.e.m, $p = 0.31$; PR: auditory-visual = −0.01 ± 0.02 $p = 0.34$, $p = 0.34$).

The lack of reward-driven effects in the PR phase could be due to a time-dependent habituation of pupil responses to reward rather than the termination of reward delivery, since the PR phase consistently occurred after the PC phase. However, we ruled out this possibility by examining the pupil responses of the first and second half of each phase (see the **Supplementary Text** and **Supplementary Figure 3**).

We next examined whether the value-driven modulation of pupil responses observed in the PC phase exhibited any correlation with the modulation of our behavioral measures. Since we observed both a modulation of accuracy (**Figure 2**) and RTs (**Figure 3**), we combined these measures into one single parameter, i.e., IE defined as the ratio of RTs of correct trials to accuracy (Vandierendonck, 2021). This parameter provides a measure of how well participants adjust their speed-accuracy trade-off. We found a strong linear relation ($\beta = -0.77$, $t_{33} = -2.59$, $p = 0.01$, **Figure 4E**) between the net effect of reward on pupil size (i.e., pupil size in high reward condition of both modalities minus pupil size in low reward of both modalities) and on IE scores. This effect indicates that a stronger value-driven pupil dilation was predictive of a stronger value-driven acceleration of visual discrimination across participants.

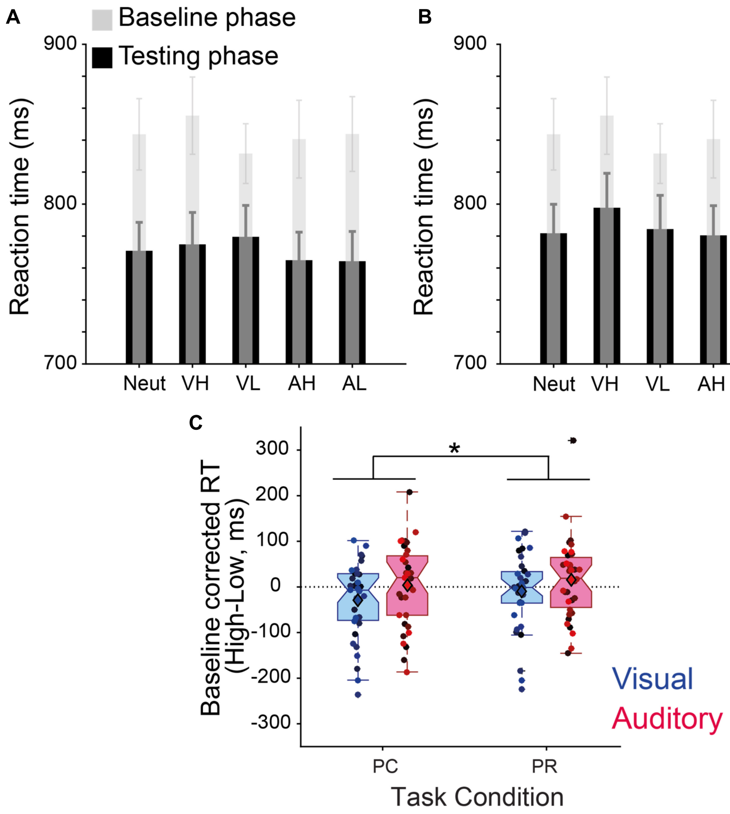


FIGURE 3
Value-driven modulation of discrimination speed. **(A)** Reaction times (RTs) (ms) of the baseline and performance-contingent reward (PC) phase. **(B)** Same as panel **(A)** for the previously associated rewards (PR) phase. The transparent gray shades represent the baseline RTs before learning the reward associations, overlaid on the test phase performance in black for each condition (neut, neutral; VH, visual high-; VL, visual low-; AH, auditory high-; and AL, auditory low-reward). **(C)** Baseline−corrected reward effect (high−low) for intra-modal (visual) and cross-modal (auditory) reward cues during the two phases. Error bars in panels **(A,B)** represent s.e.m., circles with different color shades in panel **(C)** correspond to the data of individual participants, and * stands for the interaction effect between reward and task phase at $p < 0.05$.

This correlation was non-existent in the PR phase ($\beta = 0.13$, $t_{33} = 0.35$, $p = 0.73$, **Figure 4F**).

Discussion

This study aimed to compare PC and previously associated (PR) reward cues from visual or auditory modality in terms of their modulatory effects on visual perception and task-evoked pupil responses. Our results showed that reward associated cues exert a persistent effect, in that once the reward associations are learned, reward cues improved the accuracy of perceptual judgments even when rewards were not delivered anymore (i.e., during the PR phase). PC cues were overall associated with larger task-evoked pupil responses indicating that they invoke more engagement with the task and higher goal-driven control. Furthermore, in contrast to PR, PC cues especially in visual modality, also sped up perceptual choices when a higher reward was expected and this effect was correlated with the value-driven modulation of pupil responses. These results suggest that despite the persistent effects of reward even when reward delivery is halted, some aspects of value-driven effects are specific to PC cues.

Previous research has provided evidence for a value-driven modulation of perception when the task-relevant features of stimuli are associated with high reward (Chelazzi et al., 2013; Pessoa, 2015; Failing and Theeuwes, 2018), an effect that also persists when the reward delivery is halted (De Tommaso et al., 2017). Conversely, it has been shown that the association of task-irrelevant stimuli with rewards inflicts a cost on performance, likely due to capturing attention away from the target and exhausting the cognitive control mechanisms (Sali et al., 2013; Anderson et al., 2014; Rusz et al., 2020; Watson et al., 2020). Although the majority of past research has focused on visual modality, recent studies have also examined the cross-modal effects of rewards (Leo and Noppeney, 2014; Pooresmaeili et al., 2014). Interestingly, the latter studies showed that cross-modal (auditory) stimuli that have been previously associated with higher rewards facilitated visual perception compared to low reward stimuli, despite being irrelevant to the task at hand.

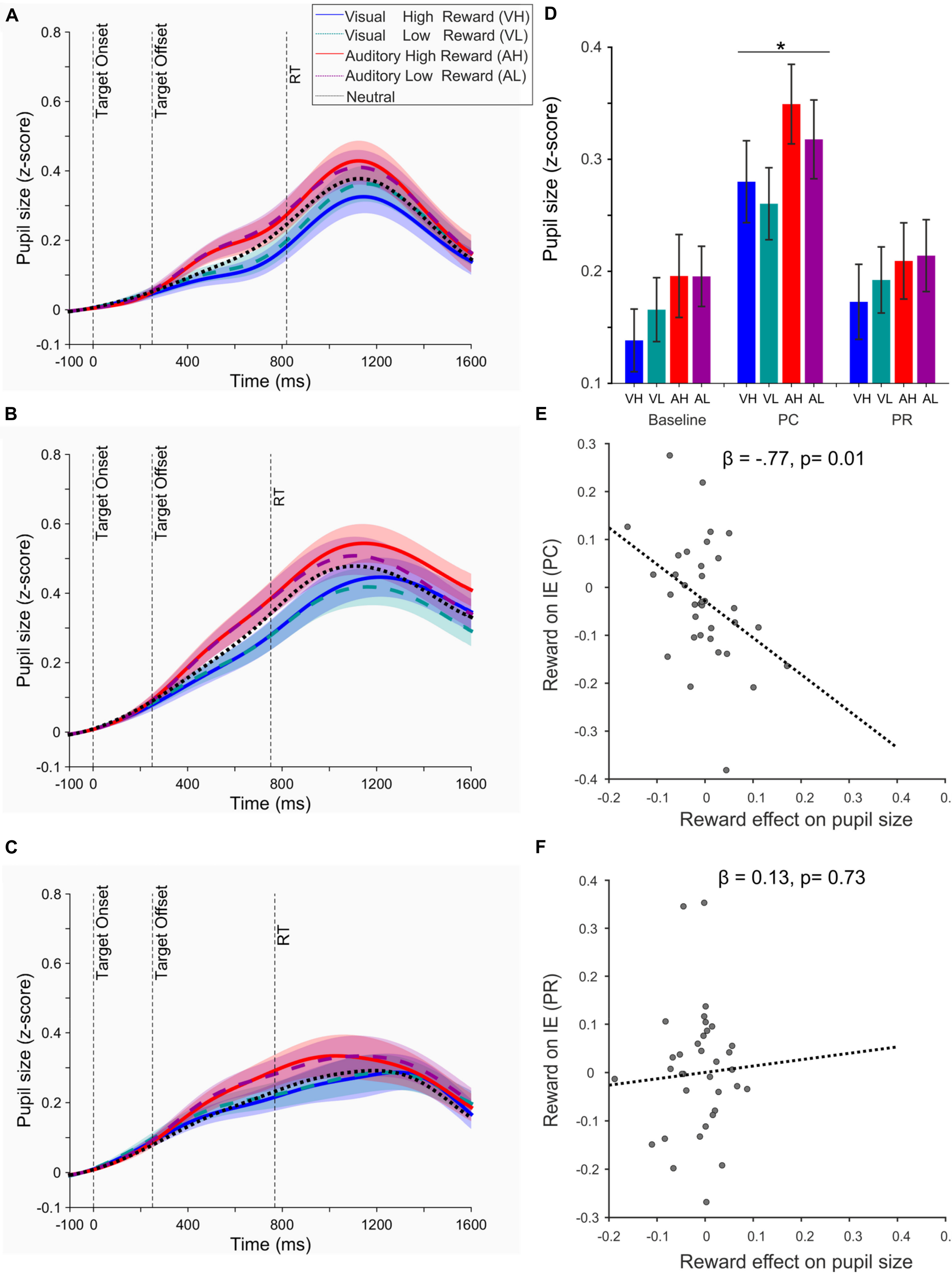


FIGURE 4

(A) Time course of pupil response for each condition during the baseline phase. **(B)** Same as panel **(A)** during the performance-contingent phase (PC). **(C)** Same as panel **(A)** during the previously associated rewards (PR). In panels **(A–C)** the vertical dashed line denoted as RT shows the mean reaction time across all conditions and across all participants. **(D)** Bar plots represent the mean task-evoked pupil size measured from the target onset until the trial end (i.e., the end of feedback phase, see **Figure 1**) for each condition (VH, visual high-; VL, visual low-; AH, auditory high-; and AL, auditory low-reward). *The effect of reward value was only significant in the PC phase at $p < 0.05$. **(E)** Relation of the value-driven modulation of pupil size (in the first 500 ms after the target onset) and inverse efficiency scores (IE) during the PC phase. **(F)** Same as panel **(E)** during the PR phase. In panels **(E,F)** regression lines are estimated based on a robust regression analysis.

These findings suggest that the value-driven increase in the salience of task-irrelevant stimuli is not necessarily associated with costs on performance. What determines whether rewards boost or impair perception in light of findings of the current study and the past research?

To understand the divergent effects observed across studies and thereby provide a unifying explanation for reward effects on perception, it is important to point to differences in the design and experimental procedures that were employed. There is a critical difference between the current study and previous studies showing that task-irrelevant reward cues captured attention away from the target and were thus associated with a cost on performance (Sali et al., 2013; Anderson et al., 2014; Rusz et al., 2020; Watson et al., 2020). In those previous studies, the majority of which employed a visual search paradigm, the target and the reward associated task-irrelevant stimuli were spatially separated. This separation might be the factor explaining the capture of attention to a different location than the target by reward cues, thereby competing with the task goal. In our study, however, both target and task-irrelevant reward cues were presented at the same spatial location, hence the capture of attention by task-irrelevant high reward cues may have spilled over to the target, increasing its representation and therefore optimizing behavior compared to low reward cues. This is in line with the findings of MacLean and Giesbrecht (2015) showing that when task-irrelevant cues were in the same location as the probed target, cues associated with higher reward magnitude improved visual search performance compared to low reward magnitude. Another related possibility is that higher reward may in fact promote perceptual grouping between the reward-associated cues and the target, as reward has been shown to interact with object-based attention (Shomstein and Johnson, 2013; Stanisor et al., 2013; Zhao et al., 2020). Therefore, in our paradigm high-reward task-irrelevant cues may have enhanced the processing of the target through a combination of space-based and object-based selection mechanisms, especially since during the PC phase these cues were predictive of the reward delivery.

The results of the current study show that PR stimuli can have long-lasting facilitatory effects on perception. However, we note that PR phase in our experiment was only tested after the PC phase, and therefore participants had a long exposure to the reward cues in a setting when they were predictive of the reward delivery when orientation discrimination task was performed correctly (i.e., the PC phase). In contrast, in our previous work (Vakhrushev et al., 2021), we tested the PR phase only after a conditioning phase which employed a different task (i.e., cue localization) than the test phase (i.e., orientation discrimination), and found that reward effects were most prominent for cross-modal cues. Together, the current results and results of our previous study indicate that the effects of reward critically depend on the training mode (Jahfari and Theeuwes, 2017; Failing and Theeuwes, 2018) and the relation between the rewarded stimuli and the task-relevant features.

Although accuracies were enhanced by high reward cues in both PC and PR phase, speed of visual discrimination was only modulated by rewards in the PC phase, especially for intra-modal cues. It is important to note that our task instructions encouraged accuracy over speed, as participants received a reward only for correct responses and independent of RT. Therefore, speeding up choices in PC occurred without an explicit instruction for speedy responses or an impact of doing so on reward magnitudes. However, by increasing the speed of choices during the PC phase for high reward cues, participants could increase their total reward rate, i.e., the amount of reward obtained in a given time for a self-paced task, a factor that has been shown to play an important role in perceptual decision making (Gold and Shadlen, 2002). When reward delivery is halted increasing the reward rate is not at stake anymore and hence in PR we did not find a speed enhancement. The motivation to increase speed in high reward PC trials, however, did not lead to a decrement in accuracy due to speed-accuracy-trade-off, suggesting that the goal-driven control mechanisms invoked by PC cues may increase the overall efficiency of perceptual choices.

Examination of pupil responses provided further evidence that PC reward cues invoke a stronger engagement of goal-driven mechanisms, as demonstrated by two key findings. Firstly, we found a stronger task-evoked pupil dilation in PC across all conditions, suggesting that in this phase participants exerted overall higher cognitive effort compared to the PR phase. Recruiting higher cognitive effort is known to increase the activity of noradrenergic neurons in Locus Coeruleus (LC) and thereby induce pupil dilation (van der Wel and van Steenbergen, 2018). Accordingly, previous studies have shown that large pupils predict the higher cognitive control required before goal-directed eye movements (Mathôt et al., 2015), reflect the higher effort required for task switching (da Silva Castanheira et al., 2021), and are indicative of the degree to which endogenous orientating of spatial attention is invoked by a task (Lasaponara et al., 2019). Importantly, the degree to which humans engage in a cognitively effortful task depends on the inherent relation between costs and benefits that ensue from performing a task (Shenhav et al., 2021) and whether the cost-benefit relations remain predictable over time (Manohar et al., 2017). In our experiment, the continuous and consistent delivery of reward upon correct performance in PC may have allowed a more direct estimation of how much rewards could compensate for the cost of extra cognitive effort, hence encouraging participants to maintain a sustained heightened level of goal-directed attention across all conditions. Secondly, in addition to the overall heightened dilation of pupils in PC phase, we found that only in this phase value-driven modulation of pupil size was significant, and this effect was predictive of the behavioral speed modulation. Modulation of pupil responses by reward value is in line with a number of previous findings (Chiew and Braver, 2013, 2014; Massar et al., 2016; Koelewijn et al., 2018; Pietrock et al., 2019; Walsh et al., 2019) and indicates that when the delivery of reward is contingent on task performance, higher reward incentives could efficiently mobilize the processing resources, and settle an efficient relationship between the speed and accuracy of choices, effects that are also reflected in the task-evoked pupil dilatation and have been reported across motor (Naber and Murphy, 2020), perceptual (Walsh et al., 2019), and cognitive (Kozunova et al., 2022) tasks. On the other hand, the lack of value-driven modulation of pupil responses for PR cues is in line with effects reported in previous studies, where reward-driven modulations of pupil size were only found during the learning of reward associations (Anderson and Yantis, 2012) but were absent during the test phase when reward-associations were implicit (Hammerschmidt et al., 2018). Taken together, these findings suggest that pupillary responses are not modulated by the mere exposure to the associative value of stimuli, but rather depend on the context in which rewards are delivered (Preuschoff et al., 2011; Cash-Padgett et al., 2018).

In the current study, the PR phase consistently occurred after the PC phase. Although our results in the PR phase could be directly compared to the previous studies that used a similar design (Vakhrushev et al., 2021), future studies would benefit from counterbalancing the task order across participants to confirm whether the results in each phase and the differences observed between PC and PR phases could be replicated. In fact, comparing our results to those reported previously (Vakhrushev et al., 2021), suggests that the reward-driven effects in the PR phase, especially for intra-modal cues, could be boosted when preceded by a phase when the delivery of rewards is PC, although this conclusion awaits future replications. In doing so, future studies may also benefit from using a larger sample size, as across experiments the effect sizes that we observed were relatively small. However, we also notice that small effect sizes could be due to the nature of the task we employed, as unlike previous studies, we used reward cues that did not carry information about the target of the visual discrimination task, a scenario when rewards and attentional requirements of the task align and larger reward driven effects are expected. Furthermore, studies on pupillometric correlates of value-driven effects can make use of paradigms in which the timing of events in each trial is tailored to the sluggish nature of pupil responses. Specifically, in our study the trial duration was relatively short (1,450–2,150 ms), which might have been insufficient to isolate the sluggish pupil modulations evoked by some of the conditions. This can be achieved by introducing a delay between the target offset and the appearance of the feedback display (see **Figure 1**) and by prolonging the intertrial intervals (ITI). Another important direction for future studies would be to further investigate which neural mechanisms give rise to the behavioral and pupillary effects that were observed here, through using neuroimaging or electrophysiological methods. This direction is important as it will allow to test whether the stronger involvement of goal-driven control during PC phase occurs through the same mechanisms that underlie attentional and reward-driven selection, namely, an enhanced engagement of fronto-parietal attentional regions (Corbetta and Shulman, 2002; Padmala and Pessoa, 2011) or changing the temporal profile of attentional control (Krebs et al., 2013). Moreover, future neuroimaging studies should investigate how the sensory modality of rewards interacts with the value-driven modulations of perception, as intra-modal and cross-modal reward effects may rely on distinct neural mechanisms (Vakhrushev et al., 2021).

In summary, our findings demonstrate a persistent effect of intra- and cross-modal rewards on visual perception. The stronger goal-driven control invoked by PC rewards and reflected in pupil responses, can additionally enhance the overall efficiency of perceptual choices by increasing the speed without sacrificing the accuracy.

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 Local Ethics Committee of the “Universitätsmedizin Göttingen” (UMG), under the proposal number 15/7/15. The patients/participants provided their written informed consent to participate in this study.

Author contributions

JA and AP conceptualized the project, interpreted the results, and wrote the first draft of the manuscript. JA conducted the experiments. AP acquired funding. All authors designed the task, analyzed the data, and revised the manuscript.

Funding

This work was supported by an ERC Starting Grant (No: 716846) to AP.

Acknowledgments

We thank Tabea Hildebrand, Jana Znaniewitz, and Sanna Peter for their help with the data collection. We also thank Prof. Melanie Wilke and Dr. Roberto Goya-Maldonado for discussions during the project presentations.

Conflict of interest

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2022.1062168/full#supplementary-material

References

Anderson, B. A., and Yantis, S. (2012). Value-driven attentional and oculomotor capture during goal-directed, unconstrained viewing. *Atten. Percept. Psychophys.* 74, 1644–1653. doi: 10.3758/s13414-012-0348-2

Anderson, B. A., Laurent, P. A., and Yantis, S. (2011). Value-driven attentional capture. *Proc. Natl. Acad. Sci. U. S. A.* 108, 10367–10371. doi: 10.1073/pnas.1104047108

Anderson, B. A., Laurent, P. A., and Yantis, S. (2014). Value-driven attentional priority signals in human basal ganglia and visual cortex. *Brain Res.* 1587, 88–96. doi: 10.1016/j.brainres.2014.08.062

Asutay, E., and Västfjäll, D. (2016). Auditory attentional selection is biased by reward cues. *Sci. Rep.* 6:36989. doi: 10.1038/srep36989

Bijleveld, E., Custers, R., and Aarts, H. (2009). The unconscious eye opener: Pupil dilation reveals strategic recruitment of resources upon presentation of subliminal reward cues. *Psychol. Sci.* 20, 1313–1315. doi: 10.1111/j.1467-9280.2009.02443.x

Botvinick, M., and Braver, T. (2015). Motivation and cognitive control: From behavior to neural mechanism. *Annu. Rev. Psychol.* 66, 83–113. doi: 10.1146/annurev-psych-010814-015044

Bucker, B., and Theeuwes, J. (2017). Pavlovian reward learning underlies value driven attentional capture. *Atten. Percept. Psychophys.* 79, 415–428. doi: 10.3758/s13414-016-1241-1

Bucker, B., and Theeuwes, J. (2018). Stimulus-driven and goal-driven effects on Pavlovian associative reward learning. *Vis. Cogn.* 26, 131–148. doi: 10.1080/13506285.2017.1399948

Camara, E., Manohar, S., and Husain, M. (2013). Past rewards capture spatial attention and action choices. *Exp. Brain Res.* 230, 291–300. doi: 10.1007/s00221-013-3654-6

Cash-Padgett, T., Azab, H., Yoo, S. B. M., and Hayden, B. Y. (2018). Opposing pupil responses to offered and anticipated reward values. *Anim. Cogn.* 21, 671–684. doi: 10.1007/s10071-018-1202-2

Chelazzi, L., Perlato, A., Santandrea, E., and Della Libera, C. (2013). Rewards teach visual selective attention. *Vision Res.* 85, 58–72. doi: 10.1016/J.VISRES.2012.12.005

Chiew, K. S., and Braver, T. S. (2013). Temporal dynamics of motivation-cognitive control interactions revealed by high-resolution pupillometry. *Front. Psychol.* 4:15. doi: 10.3389/fpsyg.2013.00015

Chiew, K. S., and Braver, T. S. (2014). Dissociable influences of reward motivation and positive emotion on cognitive control. *Cogn. Affect. Behav. Neurosci.* 14, 509–529. doi: 10.3758/s13415-014-0280-0

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755

Cornelissen, F. W., Peters, E. M., and Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behav. Res. Methods Instrum. Comput.* 34, 613–617. doi: 10.3758/BF03195489

da Silva Castanheira, K., LoParco, S., and Otto, A. R. (2021). Task-evoked pupillary responses track effort exertion: Evidence from task-switching. *Cogn. Affect. Behav. Neurosci.* 21, 592–606. doi: 10.3758/s13415-020-00843-z

De Tommaso, M., and Turatto, M. (2021). Testing reward-cue attentional salience: Attainment and dynamic changes. *Br. J. Psychol.* 113, 396–411. doi: 10.1111/bjop.12537

De Tommaso, M., Mastropasqua, T., and Turatto, M. (2017). The salience of a reward cue can outlast reward devaluation. *Behav. Neurosci.* 131, 226–234. doi: 10.1037/bne0000193

Delgado, M. R. (2007). Reward-related responses in the human striatum. *Ann. N. Y. Acad. Sci.* 1104, 70–88. doi: 10.1196/annals.1390.002

Engelmann, J. B., and Pessoa, L. (2007). Motivation Sharpens Exogenous Spatial Attention. *Emotion* 7, 668–674. doi: 10.1037/1528-3542.7.3.668

Failing, M. F., and Theeuwes, J. (2015). Nonspatial attentional capture by previously rewarded scene semantics. *Vis. Cogn.* 23, 82–104. doi: 10.1080/13506285.2014.990546

Failing, M., and Theeuwes, J. (2018). *Selection History: How Reward Modulates Selectivity of Visual Attention.* New York, NY: Springer, doi: 10.3758/s13423-017-1380-y

Gold, J. I., and Shadlen, M. N. (2002). Banburismus and the brain: Decoding the relationship between sensory stimuli, decisions, and reward. *Neuron* 36, 299–308. doi: 10.1016/S0896-6273(02)00971-6

Gong, M., Jia, K., and Li, S. (2017). Perceptual competition promotes suppression of reward salience in behavioral selection and neural representation. *J. Neurosci.* 37, 6242–6252. doi: 10.1523/JNEUROSCI.0217-17.2017

Hammerschmidt, W., Kagan, I., Kulke, L., and Schacht, A. (2018). Implicit reward associations impact face processing: Time-resolved evidence from event-related brain potentials and pupil dilations. *Neuroimage* 179, 557–569. doi: 10.1016/j.neuroimage.2018.06.055

Hickey, C., and van Zoest, W. (2012). Reward creates oculomotor salience. *Curr. Biol.* 22, R219–R220. doi: 10.1016/j.cub.2012.02.007

Itthipuripat, S., Vo, V. A., Sprague, T. C., and Serences, J. T. (2019). Value-driven attentional capture enhances distractor representations in early visual cortex. *PLoS Biol.* 17:e3000186. doi: 10.1371/journal.pbio.3000186

Jahfari, S., and Theeuwes, J. (2017). Sensitivity to value-driven attention is predicted by how we learn from value. *Psychon. Bull. Rev.* 24, 408–415. doi: 10.3758/s13423-016-1106-6

Koelewijn, T., Zekveld, A. A., Lunner, T., and Kramer, S. E. (2018). The effect of reward on listening effort as reflected by the pupil dilation response. *Hear. Res.* 367, 106–112. doi: 10.1016/j.heares.2018.07.011

Kozunova, G. L., Sayfulina, K. E., Prokofyev, A. O., Medvedev, V. A., Rytikova, A. M., Stroganova, T. A., et al. (2022). Pupil dilation and response slowing distinguish deliberate explorative choices in the probabilistic learning task. *Cogn. Affect. Behav. Neurosci.* 22, 1108–1129. doi: 10.3758/s13415-022-00996-z

Krebs, R. M., Boehler, C. N., Appelbaum, L. G., and Woldorff, M. G. (2013). Reward Associations Reduce Behavioral Interference by Changing the Temporal Dynamics of Conflict Processing. *PLoS One* 8:e53894. doi: 10.1371/journal.pone.0053894

Lasaponara, S., Fortunato, G., Dragone, A., Pellegrino, M., Marson, F., Silvetti, M., et al. (2019). Expectancy modulates pupil size both during endogenous orienting and during re-orienting of spatial attention: A study with isoluminant stimuli. *Eur. J. Neurosci.* 50, 2893–2904. doi: 10.1111/ejn.14391

Leo, F., and Noppeney, U. (2014). Conditioned sounds enhance visual processing. *PLoS One* 9:e106860. doi: 10.1371/journal.pone.0106860

MacLean, M. H., and Giesbrecht, B. (2015). Irrelevant reward and selection histories have different influences on task-relevant attentional selection. *Atten. Percept. Psychophys.* 77, 1515–1528. doi: 10.3758/s13414-015-0851-3

Macmillan, N. A., and Creelman, C. D. (1991). *Detection Theory: A User's Guide*. New York, NY: Cambridge University Press.

Manohar, S. G., Finzi, R. D., Drew, D., and Husain, M. (2017). Distinct Motivational Effects of Contingent and Noncontingent Rewards. *Psychol. Sci.* 28, 1016–1026. doi: 10.1177/0956797617693326

Massar, S. A. A., Lim, J., Sasmita, K., and Chee, M. W. L. (2016). Rewards boost sustained attention through higher effort: A value-based decision making approach. *Biol. Psychol.* 120, 21–27. doi: 10.1016/j.biopsycho.2016.07.019

Mathôt, S., Siebold, A., Donk, M., and Vitu, F. (2015). Large pupils predict goal-driven eye movements. *J. Exp. Psychol. Gen.* 144, 513–521. doi: 10.1037/a0039168

Maunsell, J. H. R. (2004). Neuronal representations of cognitive state: Reward or attention? *Trends Cogn. Sci.* 8, 261–265. doi: 10.1016/J.TICS.2004.04.003

Milstein, D. M., and Dorris, M. C. (2007). The influence of expected value on saccadic preparation. *J. Neurosci.* 27, 4810–4818. doi: 10.1523/JNEUROSCI.0577-07.2007

Naber, M., and Murphy, P. (2020). Pupillometric investigation into the speed-accuracy trade-off in a visuo-motor aiming task. *Psychophysiology* 57:e13499. doi: 10.1111/psyp.13499

Padmala, S., and Pessoa, L. (2011). Reward reduces conflict by enhancing attentional control and biasing visual cortical processing. *J. Cogn. Neurosci.* 23, 3419–3432. doi: 10.1162/jocn_a_00011

Pessoa, L. (2015). Multiple influences of reward on perception and attention. *Vis. Cogn.* 23, 272–290. doi: 10.1080/13506285.2014.974729

Pessoa, L., and Engelmann, J. B. (2010). Embedding reward signals into perception and cognition. *Front. Neurosci.* 4:17. doi: 10.3389/fnins.2010.00017

Pietrock, C., Ebrahimi, C., Katthagen, T. M., Koch, S. P., Heinz, A., Rothkirch, M., et al. (2019). Pupil dilation as an implicit measure of appetitive Pavlovian learning. *Psychophysiology* 56:e13463. doi: 10.1111/psyp.13463

Pleger, B., Blankenburg, F., Ruff, C. C., Driver, J., and Dolan, R. J. (2008). Reward facilitates tactile judgments and modulates hemodynamic responses in human primary somatosensory cortex. *J. Neurosci.* 28, 8161–8168. doi: 10.1523/JNEUROSCI.1093-08.2008

Pooresmaeili, A., FitzGerald, T. H. B., Bach, D. R., Toelch, U., Ostendorf, F., and Dolan, R. J. (2014). Cross-modal effects of value on perceptual acuity and stimulus encoding. *Proc. Natl. Acad. Sci. U. S. A.* 111, 15244–15249. doi: 10.1073/pnas.1408873111

Preuschoff, K., 't Hart, B. M., and Einhäuser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. *Front. Neurosci.* 5:115. doi: 10.3389/fnins.2011.00115

Qin, N., Xue, J., Chen, C., and Zhang, M. (2020). The Bright and Dark Sides of Performance-Dependent Monetary Rewards: Evidence From Visual Perception Tasks. *Cogn. Sci.* 44:e12825. doi: 10.1111/cogs.12825

Rusz, D., Le Pelley, M. E., Kompier, M. A. J., Mait, L., and Bijleveld, E. (2020). Reward-Driven distraction: A meta-analysis. *Psychol. Bull.* 146, 872–899. doi: 10.1037/bul0000296

Sali, A. W., Anderson, B. A., and Yantis, S. (2013). The Role of Predictable and Unpredictable Reward in the Control of Attention. *J. Vis.* 13, 892–892. doi: 10.1167/13.9.892

Schneider, M., Leuchs, L., Czisch, M., Sämann, P. G., and Spoormaker, V. I. (2018). Disentangling reward anticipation with simultaneous pupillometry / fMRI. *Neuroimage* 178, 11–22. doi: 10.1016/j.neuroimage.2018.04.078

Schultz, W. (2015). Neuronal reward and decision signals: From theories to data. *Physiol. Rev.* 95, 853–951. doi: 10.1152/physrev.00023.2014

Serences, J. T. (2008). Value-based modulations in human visual cortex. *Neuron* 60, 1169–1181. doi: 10.1016/j.neuron.2008.10.051

Seriès, P., and Seitz, A. R. (2013). Learning what to expect (in visual perception). *Front. Hum. Neurosci.* 7:668. doi: 10.3389/fnhum.2013.00668

Shenhav, A., Prater Fahey, M., and Grahek, I. (2021). Decomposing the Motivation to Exert Mental Effort. *Curr. Dir. Psychol. Sci.* 30, 307–314. doi: 10.1177/09637214211009510

Shomstein, S., and Johnson, J. (2013). Shaping Attention With Reward: Effects of Reward on Space- and Object-Based Selection. *Psychol. Sci.* 24, 2369–2378. doi: 10.1177/0956797613490743

Shuler, M. G., and Bear, M. F. (2006) Reward timing in the primary visual cortex. *Science* 311, 1606–1609.

Stanisor, L., van der Togt, C., Pennartz, C. M. A., and Roelfsema, P. R. (2013). A unified selection signal for attention and reward in primary visual cortex. *Proc. Natl. Acad. Sci. U S A.* 110, 9136–9141. doi: 10.1073/pnas.1300117110

Tankelevitch, L., Spaak, E., Rushworth, M. F. S., and Stokes, M. G. (2020). Previously Reward-Associated Stimuli Capture Spatial Attention in the Absence of Changes in the Corresponding Sensory Representations as Measured with MEG. *J. Neurosci.* 40, 5033–5050. doi: 10.1523/JNEUROSCI.1172-19.2020

Vakhrushev, R., Cheng, F., Schacht, A., and Pooresmaeili, A. (2021). Differential effects of intra-modal and cross-modal reward value on visual perception: ERP evidence. *BioRxiv* [Preprint]. 2021.09.29.462374. doi: 10.1101/2021.09.29.462374

van der Wel, P., and van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychon. Bull. Rev.* 25, 2005–2015. doi: 10.3758/s13423-018-1432-y

Vandierendonck, A. (2021). On the Utility of Integrated Speed-Accuracy Measures when Speed-Accuracy Trade-off is Present. *J. Cogn.* 4:22. doi: 10.5334/joc.154

Wallis, J. D. (2007). Orbitofrontal cortex and its contribution to decision-making. *Annu. Rev. Neurosci.* 30, 31–56. doi: 10.1146/annurev.neuro.30.051606.094334

Walsh, A. T., Carmel, D., and Grimshaw, G. M. (2019). Reward elicits cognitive control over emotional distraction: Evidence from pupillometry. *Cogn. Affect. Behav. Neurosci.* 19, 537–554. doi: 10.3758/s13415-018-00669-w

Watson, A. B., and Pelli, D. G. (1983). Quest: A Bayesian adaptive psychometric method. *Percept. Psychophys.* 33, 113–120. doi: 10.3758/BF03202828

Watson, P., Pearson, D., Theeuwes, J., Most, S. B., and Le Pelley, M. E. (2020). Delayed disengagement of attention from distractors signalling reward. *Cognition* 195:104125. doi: 10.1016/j.cognition.2019.104125

Weil, R. S., Furl, N., Ruff, C. C., Symmonds, M., Flandin, G., Dolan, R. J., et al. (2010) Rewarding feedback after correct visual discriminations has both general and specific influences on visual cortex. *J. Neurophysiol.* 104, 1746–1757.

Yantis, S., Anderson, B. A., Wampler, E. K., and Laurent, P. A. (2012). "Reward and Attentional Control in Visual Search," in *Nebraska Symposium on Motivation. Nebraska Symposium on Motivation*, eds M. D. Dodd and J. H. Flowers (Lincoln, NE: University of Nebraska Press), 91–116. doi: 10.1007/978-1-4614-4794-8_5

Zhao, J., Song, F., Zhou, S., Hu, S., Liu, D., Wang, Y., et al. (2020). The impact of monetary stimuli on object-based attention. *Br. J. Psychol.* 111, 460–472. doi: 10.1111/bjop.12418

frontiers | Frontiers in Human Neuroscience

TYPE Original Research
PUBLISHED 04 January 2023
DOI 10.3389/fnhum.2022.1027335

OPEN ACCESS

EDITED BY
Alessia Tonelli,
Italian Institute of Technology (IIT), Italy

REVIEWED BY
John Magnotti,
University of Pennsylvania,
United States
Marc Mathijs Van Wanrooij,
Radboud University, Netherlands

*CORRESPONDENCE
Leslie D. Kwakye
✉ lkwakye@oberlin.edu

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

SPECIALTY SECTION
This article was submitted to
Sensory Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 24 August 2022
ACCEPTED 05 December 2022
PUBLISHED 04 January 2023

CITATION
Fisher VL, Dean CL, Nave CS, Parkins EV, Kerkhoff WG and Kwakye LD (2023) Increases in sensory noise predict attentional disruptions to audiovisual speech perception.
Front. Hum. Neurosci. 16:1027335.
doi: 10.3389/fnhum.2022.1027335

Increases in sensory noise predict attentional disruptions to audiovisual speech perception

Victoria L. Fisher[1,2]†, Cassandra L. Dean[1,3]†, Claire S. Nave[1], Emma V. Parkins[1,4], Willa G. Kerkhoff[1,5] and Leslie D. Kwakye[1]*

[1]Department of Neuroscience, Oberlin College, Oberlin, OH, United States, [2]Yale University School of Medicine and the Connecticut Mental Health Center, New Haven, CT, United States, [3]Roche/Genentech Neurodevelopment & Psychiatry Teams Product Development, Neuroscience, South San Francisco, CA, United States, [4]Neuroscience Graduate Program, University of Cincinnati, Cincinnati, OH, United States, [5]Department of Neurobiology, University of Pittsburgh, Pittsburgh, PA, United States

We receive information about the world around us from multiple senses which combine in a process known as multisensory integration. Multisensory integration has been shown to be dependent on attention; however, the neural mechanisms underlying this effect are poorly understood. The current study investigates whether changes in sensory noise explain the effect of attention on multisensory integration and whether attentional modulations to multisensory integration occur via modality-specific mechanisms. A task based on the McGurk Illusion was used to measure multisensory integration while attention was manipulated via a concurrent auditory or visual task. Sensory noise was measured within modality based on variability in unisensory performance and was used to predict attentional changes to McGurk perception. Consistent with previous studies, reports of the McGurk illusion decreased when accompanied with a secondary task; however, this effect was stronger for the secondary visual (as opposed to auditory) task. While auditory noise was not influenced by either secondary task, visual noise increased with the addition of the secondary visual task specifically. Interestingly, visual noise accounted for significant variability in attentional disruptions to the McGurk illusion. Overall, these results strongly suggest that sensory noise may underlie attentional alterations to multisensory integration in a modality-specific manner. Future studies are needed to determine whether this finding generalizes to other types of multisensory integration and attentional manipulations. This line of research may inform future studies of attentional alterations to sensory processing in neurological disorders, such as Schizophrenia, Autism, and ADHD.

KEYWORDS

multisensory integration (MSI), attention, dual task, McGurk effect, perceptual load, audiovisual speech, sensory noise, neural mechanisms

1. Introduction

The interactions between top-down cognitive processes and multisensory integration have been heavily investigated and shown to be intricate and multidirectional (Talsma et al., 2010; Cascio et al., 2016; Stevenson et al., 2017). Previous research using different methods to manipulate attention and measure multisensory integration has demonstrated that multisensory integration is lessened under high attentional demand and relies on the distribution of attention to all stimuli being integrated (Alsius et al., 2005, 2007; Talsma et al., 2007; Mozolic et al., 2008; Koelewijn et al., 2010; Tang et al., 2016; Gibney et al., 2017). Studies investigating the time point(s) during which attentional alterations influence multisensory processing have identified both early and late attentional effects (Talsma and Woldorff, 2005; Talsma et al., 2007; Mishra et al., 2010). Additionally, multiple areas such as the Superior Temporal Sulcus (STS), Superior Temporal Gyrus (STG), and extrastriate cortex have been identified as cortical loci of attentional changes to multisensory processing (Mishra and Gazzaley, 2012; Morís Fernández et al., 2015). Collectively, these studies suggest that attention alters multisensory processing at multiple time points and cortical sites throughout the sensory processing hierarchy.

The precise mechanisms by which attention alters multisensory integration remain unknown. Multisensory percepts are built through hierarchical processing within sensory systems, coherent activity across multiple cortical sites, and convergence onto heteromodal areas (for an extensive review see Engel et al., 2012). Alterations in attention may primarily disrupt multisensory integration by interfering with integrative processes such as synchronous oscillatory activity across cortical areas or processing of multisensory information within heteromodal areas (Senkowski et al., 2005; Schroeder et al., 2008; Koelewijn et al., 2010; Al-Aidroos et al., 2012; Friese et al., 2016). Attention and oscillatory synchrony have been shown to interact in a number of studies (Lakatos et al., 2008; Gomez-Ramirez et al., 2011; Keil et al., 2016), thus strengthening the possibility of this potential mechanism. Although there is convincing evidence for attentional changes to integrative processes, there is a strong likelihood that disruptions in unisensory processing may explain, in part, attentional alterations in multisensory integration. An extensive research literature clearly demonstrates that attention influences unisensory processing within each sensory modality (Woldorff et al., 1993; Mangun, 1995; Driver, 2001; Pessoa et al., 2003; Mitchell et al., 2007; Okamoto et al., 2007; Ling et al., 2009). Additionally, attention has been shown to improve the neural encoding of auditory speech in lower-order areas and to selectively encode attended speech in higher-order areas (Zion Golumbic E. et al., 2013; Zion Golumbic E. M. et al., 2013). Alterations in the reliability of unisensory components of multisensory stimuli have been clearly demonstrated to alter patterns of multisensory integration such that the brain more heavily weighs input from the modality providing the clearest information (Deneve and Pouget, 2004; Bobrowski et al., 2009; Burns and Blohm, 2010; Magnotti et al., 2013, 2020; Magnotti and Beauchamp, 2015, 2017; Noel et al., 2018a). Thus, disruptions in attention may result in increased neural variability during stimulus encoding (sensory noise) causing degraded unisensory representations to be integrated into altered multisensory perceptions. Few studies have directly assessed the impact of attention on sensory noise and multisensory integration (Schwartz et al., 2010; Odegaard et al., 2016); thus, more exploration is needed to determine whether attentional influences on multisensory integration may be explained by increases in sensory noise.

Psychophysical tasks utilizing multisensory illusions may be able to determine whether attentional alterations in multisensory integration are mediated by disruptions in modality-specific processing. Multisensory illusions which result from discrepancies in information across modalities are ideally suited for this type of experimental design because the strength of the illusion can be altered by changing the reliability of the component unisensory stimuli and these effects can be modeled by measuring the ratio of visual and auditory sensory noise (Körding et al., 2007; Magnotti and Beauchamp, 2017). The McGurk effect is a well-known illusion that has been used to study multisensory speech perception (Mcgurk and Macdonald, 1976) and the effects of attention on audiovisual speech integration. The strength of the McGurk effect has consistently been shown to decrease with increasing perceptual load in dual-task studies (Paré et al., 2003; Alsius et al., 2005, 2007, 2014; Soto-Faraco and Alsius, 2009; Gibney et al., 2017). Because audiovisual speech can be understood through its unisensory components and requires extensive processing of the speech signal prior to integration (Zion Golumbic E. et al., 2013; Zion Golumbic E. M. et al., 2013), there is a strong likelihood that attentional alterations in audiovisual speech integration may be explained by disruptions to the unisensory processing of speech information. Specifically, disruptions in the encoding of visual speech components would be expected to weaken the McGurk Effect while disruptions in the encoding of auditory speech components would strengthen the McGurk Effect.

In this study, we investigate attentional influences on early auditory and visual processing by examining modality-specific attentional changes to sensory noise. In two separate experiments, participants completed a McGurk task that included unisensory and congruent multisensory trials while concurrently completing a secondary auditory or visual task. Sensory noise was calculated from the variability in participants' unisensory responses separately for the auditory and visual modalities. Multiple regression analysis (MRA) was then used to determine the impact of visual noise, auditory noise, and distractor modality on McGurk reports at baseline and changes in McGurk reports with increasing perceptual load. We predicted that increases in perceptual load would lead to decreases in the McGurk effect and increases in sensory noise within the same modality as the distractor. Additionally, we predicted that changes in McGurk reports with increasing load would be best predicted by changes in visual noise (as compared to changes in auditory noise).

2. Materials and methods

2.1. Participants

A total of 172 (120 Females, 18–44 years of age, mean age of 22) typically developing adults completed this study. 57 (38 Females, 18–36 years of age, mean age of 22) participants completed trials with auditory distractors and 138 (82 Females, 18–44 years of age, mean age of 22) participants completed trials with visual distractors. Data from some participants overlaps with data previously published in Gibney et al. (2017). Twenty-three (23) participants completed both experiments in separate sessions. Participants were excluded from final analysis if they did not complete at least four repetitions of every trial type (45) or did not have a total accuracy of at least 60% on the distractor task for the high load condition (12). Thus, 115 participants were included in the final analysis. Participants reported normal or corrected-to-normal hearing and vision and no prior history of seizures. Participants gave written informed consent and were compensated for their time. Study procedures were conducted under the guidelines of Helsinki and approved by the Oberlin College Institutional Review Board.

2.2. Experimental design overview

We employed a dual-task design to determine the effects of attention within a specific sensory modality on McGurk perceptions and on sensory noise within each modality. Similar dual task designs have been shown to reduce attentional capacity (Lavie et al., 2003; Stolte et al., 2014; Bonato et al., 2015). Participants completed a primary McGurk task concurrently with a secondary visual or auditory distractor task for which the level of visual or auditory perceptual load was modulated. Full methodology for both the primary McGurk task as well as the secondary distractor tasks has been previously published in Dean et al. (2017) and Gibney et al. (2017); however, we provide a brief overview of all tasks here. All study procedures were completed in a dimly lit, sound-attenuated room. Participants were monitored via closed-circuit cameras for safety and to ensure on-task behavior. All visual stimuli were presented on a 24" Asus VG 248 LCD monitor at a screen resolution of 1,920°×°1,080 with a refresh rate of 144 Hz at a viewing distance of 50 cm from the participant. All auditory stimuli were presented from Dual LU43PB speakers which were powered by a Lepas LP-2020AC 2-Ch digital amplifier and were located to the right and left of the participant. SuperLab 4.5 software was used for stimulus presentation and participant response collection. Participants indicated their responses on a Cedrus RB-834 response box, and responses were saved to a txt file.

2.3. McGurk task

Participants were presented with videos of a woman speaking one of four syllables "ba" (/ba/), "ga" (/ga/), "da" (/da/), or "tha" (/tha/, voiceless) (**Figure 1A**). Trials were either unisensory (visual-only; auditory-only) or multisensory (congruent; incongruent illusory; incongruent non-illusory). In unisensory trials, participants were presented with either the visual (visual-only) or auditory (auditory-only) components of the video for each syllable. Multisensory videos had both an auditory and a visual component and were either congruent (e.g., visual "ba" auditory "ba"), incongruent non-illusory (visual "ba" auditory "ga"), or incongruent illusory (visual "ga" auditory "ba"). Participants responded to the prompt, "What did she say?" by pushing one of four buttons labeled "ba," "ga," "da," or "tha." Although eye movements were not monitored, participants were explicitly instructed to maintain their gaze on the speaker's mouth throughout the duration of the study. Each unisensory syllable was repeated 8 times for a total of 32 visual-only and 32 auditory-only trials. Each congruent multisensory syllable was repeated 8 times for a total of 32 total congruent multisensory trials. Lastly, there were 16 illusory incongruent and 16 non-illusory incongruent trials.

2.4. Secondary visual distractor task

Rapid serial visual presentation (RSVP) stimuli of white letters, yellow letters, and white numbers presented continuously below the McGurk videos (**Figure 1A**). Each letter and number in the RSVP stream was presented for 100 ms with 20 ms between letters and numbers. The visual distractor task included four condition types: distractor free (DF), no perceptual load (NL), low perceptual load (LL), and high perceptual load (HL). During distractor-free blocks, no visual or auditory distractors were presented; thus, participants completed the McGurk task in isolation. When the RSVP stream was presented concurrently with the McGurk task, participants were asked to either ignore it (NL), detect infrequent yellow letters (LL), or detect infrequent white numbers (HL). There was a 50% chance that the target would be present in each trial. After each presentation, participants were asked to respond first to the McGurk task then report whether they observed a target within the RSVP stream with a "yes" or "no" button press. Each load condition was completed in a separate block, and the order of blocks was randomized and counterbalanced across participants. Participants completed all perceptual load blocks in one session.

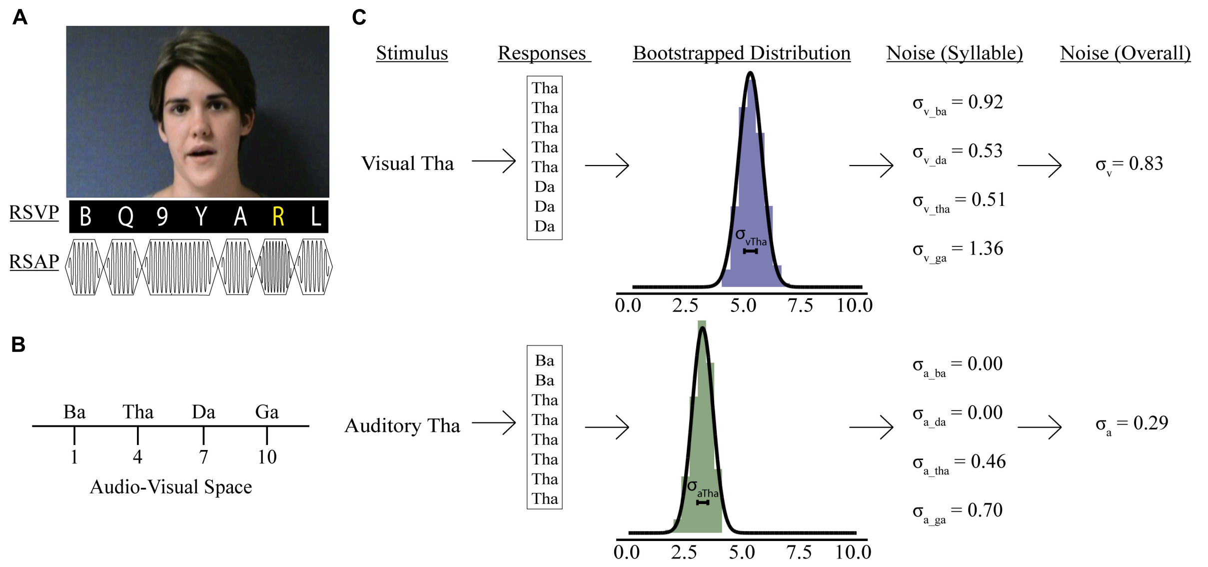


FIGURE 1

Psychophysics tasks and sensory noise calculations. **(A)** Participants watched videos of a woman speaking one of four syllables, after which they reported if she said: "ba," "ga," "da," or "tha." Rapid serial visual presentation (RSVP) or rapid serial auditory presentation (RSAP) stimuli accompanied speech videos during no load (NL), low load (LL), and high load (HL) blocks. For the visual distractor task, participants detected a yellow letter (LL) or a white number (HL). For the auditory distractor task, participants detected a high-pitched tone (LL) or a long-duration tone (HL). Identifiable human image used with permission. **(B)** Mapping of possible responses in representative audio-visual space. Panel **(C)** shows sensory noise calculations for an example participant. Sensory noise was calculated for each participant using responses from visual (top) and auditory (bottom) only trials. Gaussian distributions of these responses were determined via bootstrapping (middle), and the standard deviation of this distribution was calculated for each syllable. The overall visual (top, last panel) and auditory (bottom, last panel) noise for each participant was calculated as the average standard deviation of all syllabi within each modality.

2.5. Secondary auditory distractor task

Stimuli consisted of rapid serial auditory presentation (RSAP) of musical notes at frequencies between 262 and 523 Hz. Each note was presented for 100 ms with 20 ms between notes (Figure 1A). As in the visual distractor task, there were four auditory perceptual load conditions: no distractors presented alongside McGurk stimuli (DF); distractor stimuli were present but not attended (NL), participants were asked to detect a tone significantly higher pitch (1,046–2,093 Hz) than the standard tones (LL); participants were asked to detect notes that were twice the duration of the standard tones (HL). For LL and HL trials, there was a 50% probability that the target would be present in the RSAP stream. After each presentation, participants first responded to the McGurk task, then selected "Yes" or "No" to indicate if they observed the target. Participants completed all perceptual load blocks in one session.

2.6. Data analysis

2.6.1. Psychophysical analyses

Responses for incongruent illusory trials on the McGurk task were divided into "visual" ("ga"), "auditory" ("ba"), and "fused" ("da" or "tha"). Percent fused reports were calculated for each participant for each perceptual load condition and distractor modality. We conducted a repeated-measures analysis of variance (RMANOVA) on percent fused reports with load (NL or HL) as a within-subject factor and distractor task modality (visual or auditory) as a between-subjects factor to determine whether increasing perceptual load affected the perception of the McGurk Illusion and whether this effect was modulated by distractor modality.

2.6.2. Sensory noise calculations

Previous models have been developed to determine sensory noise (Magnotti and Beauchamp, 2015, 2017). However, these models do not account for visual and auditory noise independently. Including visual and auditory noise independently permits investigations into how distractors impact precision of information available when forming McGurk percepts, which may be important for understanding attentional influences on multisensory integration. We assessed sensory noise in both modalities using variability in responses to unisensory visual and auditory presentations. Previous studies determined that the encoding of auditory and visual cues follow separate Gaussian distributions and that the variance of that distribution reflects sensory noise (Ma et al., 2009; Magnotti and Beauchamp, 2017). Responses to visual and auditory-only trials were used to estimate sensory noise separately for each experimental condition: syllable presented ("ba," "tha," "da," "ga"), distractor modality (auditory or visual), and perceptual load (DF, NL, or HL). Each response was assigned a value reflecting the reported syllable's relative location in audiovisual perceptual space (**Figure 1B**; Ma et al., 2009; Olasagasti et al., 2015; Magnotti and Beauchamp, 2017; Lalonde and Werner, 2019). In line with previous work, fused reports were placed in the middle of "ba" and "ga" (Magnotti and Beauchamp, 2017). However, our study design permitted two options "da" and "tha" for fused responses. To account for differences in between the two syllables we adapted a 10-point scale. This would permit us to separate "tha" and "da," to accommodate previous findings that "tha" is more similar to "ba," while "da" is more similar to "ga" (Lalonde and Werner, 2019). Further, Lalonde and Werner identified multiple consonant-groups separating each syllable, thus a 10-point scale would reflect distance in audiovisual space between each syllable.

We bootstrapped 10,000 samples for each participant's response to each syllable presented during auditory- and visual-only trials (**Figure 1C**, Stein et al., 2009). We averaged each syllable's overall visual (σ_{Vis}) and auditory (σ_{Aud}) noise for each condition by taking the average sensory noise for all syllables presented during visual or auditory-only trials. Finally, we calculated combined sensory noise to account for both visual and auditory noise. We used the equation: $\sigma_{Combined} = \frac{\sigma_{Vis} - \sigma_{Aud}}{\sigma_{Vis} + \sigma_{Aud}}$, which is based on calculations from maximum likelihood estimate models (Ernst and Banks, 2002) and comparable to models using auditory/visual noise ratio (Magnotti and Beauchamp, 2017; Magnotti et al., 2018). This produces a distribution of combined sensory noise values between 1 and -1, with values >0 indicating that visual noise is greater.

2.6.3. Multiple regression modeling

We developed two multiple linear regression models to determine the effect of sensory noise on McGurk perceptions. We chose to use linear regression because to investigate the roles of attention and sensory noise on the likelihood of perceiving the McGurk effect. Additionally, relevant factors used in the analyses showed significant linear relationships with our dependent factors. The first model investigated factors contributing to McGurk responses at baseline, and the second investigated changes in McGurk responses with increasing perceptual load. All testing and model assessments were carried out in SPSS. First, preliminary model fitting was conducted on data from individuals excluded ($n = 57$) due to poor distractor task performance and lack of unisensory data to explore the relationship between baseline McGurk values and multiple possible predictor variables. These variables included visual noise, auditory noise, distractor modality, accuracy on auditory and visual distractor tasks, and interaction terms. Preliminary results suggested that visual noise, auditory noise, and the combination of the two could be predictive of McGurk responses. After determining potential predictors from excluded data, we then determined whether McGurk responses at baseline (distractor-free condition) correlated with each sensory noise measure (visual, auditory, and combined) to construct the final multiple regression model. Importantly, this baseline regression model allowed us to better contextualize our results and our novel method of estimating sensory noise within modality in the context of previous studies which also relate sensory noise to measures of multisensory integration.

Our second multiple regression analysis modeled the change in McGurk perception from NL to HL (ΔMcGurk $=$ HL McGurk reports $-$ NL McGurk reports). To determine which predictive variables to include, we performed an RMANOVA with visual noise, auditory noise, and combined noise as dependent variables with load (NL and HL) as a within-subjects factor and distractor modality as a between-subjects variable. The variables that were significantly predicted by load were included in a single-step multiple regression model of ΔMcGurk: distractor modality, change in visual noise, and baseline McGurk values. Notably, changes in auditory noise and combined noise were excluded because neither these variables nor their interaction with distractor modality were significantly predicted by load nor did they correlate with changes in McGurk reports across load.

3. Results

Participants completed a McGurk detection task to assess their integration of speech stimuli. This task was completed alone (DF) or in addition to a secondary distractor task at various perceptual loads (NL and HL). Participants were separated by which distractor modality (auditory or visual) was presented during the dual-task conditions.

3.1. Attentional alterations to McGurk perception

To assess baseline levels of multisensory integration, percent fused responses ("da" or "tha") were calculated for illusory trials (auditory "ba" and visual "ga") during the distractor-free block (**Figure 2**). Independent t-tests revealed significant differences in mean baseline illusory percepts between the auditory distractor group (percent fused $= 41.05$) and visual distractor group (percent fused $= 68.11$; $t_{105} = 4.54$, $p = 1.50 \times 10^{-5}$, Cohen's $d = 0.724$). These differences were confirmed with bootstrapped (95% CI: 4.45–32.04, $p = 0.015$), non-parametric ($\mathrm{U}_{N,AudDist:134;\ N,VisDist:58} = 2{,}191$, $z = -4.85$, $p = 1.26 \times 10^{-6}$) and Bayesian ($t_{190} = 4.81$, $p = 7.4 \times 10^{-6}$, $BF = 0.00$) sample comparisons. Because the distractor-free block was identical for the visual and auditory distractor studies and was most often completed after a NL, LL, or HL block, these results may indicate

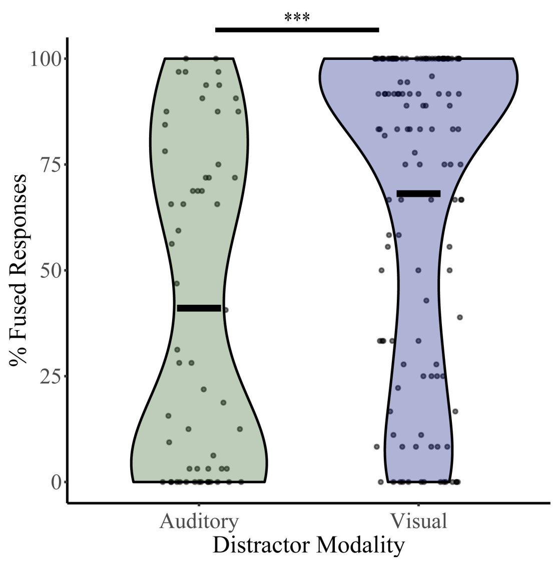


FIGURE 2
McGurk fused reports for distractor free blocks. The percent of fused reports ("da" or "tha") during distractor free blocks is shown for each participant for the visual distractor and auditory distractor groups. Horizontal black bars indicate group averages, and violin plots display the distribution of percent fused reports for each task. ***Indicates $p < 0.001$.

that McGurk perception is affected by the modality of distractors within the context of the entire task.

To assess how McGurk perception changes with increasing perceptual load, we calculated fused responses during no-load and high load blocks (Figure 3) for both the auditory distractor group (NL %fused = 45.90, HL %fused = 37.80) and visual distractor group (NL %fused = 60.86, HL %fused = 33.68%). A two-way RMANOVA with fused responses as the dependent factor, perceptual load as a within-subjects factor, and distractor modality as a between-subjects factor revealed a main effect of perceptual load ($F_{1,133} = 48.36$, $p = 1.45 \times 10^{-10}$, partial $\eta^2 = 0.267$) and an interaction between load and distractor modality ($F_{1,133} = 14.15$, $p = 2.52 \times 10^{-4}$, partial $\eta^2 = 0.096$). We confirmed these findings with *post hoc* two-sample comparisons. These indicate significant changes in McGurk responses from No Load to High Load with visual distractors ($t_{86} = 8.36$, $p = 9.75 \times 10^{-13}$, Cohen's d: 0.90; Bootstrapped 95% CI: 20.76–33.72, $p = 2.00 \times 10^{-4}$; $W = 114.5$, $z = -6.67$, $p = 2.59 \times 10^{-11}$; $BF = 0.00$). Parametric assessments illustrated a significant change in McGurk responses between auditory No Load to High Load ($t_{47} = 2.35$, $p = 0.02$, Cohen's d: 0.34; Bootstrapped 95% CI: 1.62–15.09, $p = 0.032$; $BF = 0.67$); however, this effect only approached significance when using non-parametric Wilcoxon comparisons ($W = 271.50$, $z = -1.86$, $p = 0.06$). Further, differences in McGurk reports from No Load to High Load conditions were dependent on distractor modality ($t_{117} = -4.03$, $p = 1.01 \times 10^{-4}$, Cohen's $d = -0.68$; Bootstrapped 95% CI: −28.35 to −9.86, $p = 2.00 \times 10^{-4}$; $U_{N,AudDist:48;N,VisDist:87} = 2{,}918$, $z = 3.83$, $p = 1.30 \times 10^{-4}$; Bayesian $t_{133} = -3.8$, $p = 2.532 \times 10^{-4}$, $BF = 0.012$). These results indicate that increasing perceptual load leads to a decrease in integration; however, visual distractors led to a greater decrease in integration than auditory distractors. Supplementary material include figures and statistics for participant distractor task accuracy (Supplementary Figure 1), unisensory and multisensory congruent trial-type accuracy (Supplementary Figure 2), and changes in McGurk reports across NL, LL, and HL (Supplementary Figure 3) for both distractor modalities.

3.2. Sensory noise

3.2.1. Baseline sensory noise

Responses on unisensory trials were used to determine auditory and visual noise values for each participant during baseline conditions (distractor free block; Figure 4). Both visual distractor group (σ_{Vis} 0.50, σ_{Aud} 0.11) and auditory distractor group (σ_{Vis} 0.54, σ_{Aud} 0.11) had lower auditory noise than visual noise. A two-way ANOVA with sensory noise as the dependent variable, noise modality as a within-subjects factor, and distractor modality as a between-subjects factor revealed a main effect of noise modality ($F_{1,190} = 450$, $p = 4.83 \times 10^{-52}$, partial $\eta^2 = 0.703$). There was no effect of distractor modality ($F_{1,190} = 1.092$, $p = 0.297$, partial $\eta^2 = 0.006$) or interaction between noise and distractor modality ($F_{1,190} = 0.948$, $p = 0.331$, partial $\eta^2 = 0.005$). *Post hoc* sample comparisons using t-tests and non-parametric assessments corroborated these findings. There were significant differences between baseline auditory and visual noise for individuals in both auditory-distractor ($t_{47} = 10.93$, $p = 1.70 \times 10^{-14}$, Cohen's d: 1.58; Bootstrapped 95% CI: 0.34–0.48, $p = 2.00 \times 10^{-4}$; $W = 24$, $z = -6.44$, $p = 1.21 \times 10^{-10}$; $t_{57} = 12.3$, $p = 0.000$, $BF = 0.00$) and visual distractor ($t_{86} = 13.78$, $p = 1.78 \times 10^{-23}$, Cohen's d: 1.48; Bootstrapped 95% CI: 1.62–15.09, $p = 0.03$; $W = 89.00$, $z = -9.85$, $p = 0.000$; Bayesian $t_{134} = 19.2$, $p = 0.000$, $BF = 0.00$) groups. These results indicate that auditory noise was significantly lower than visual noise regardless of the distractor modality for the task.

3.2.2. Change in sensory noise

Next, we investigated whether perceptual load increased sensory noise and whether this effect was dependent on distractor or noise modality (Figure 5). For the auditory distractor group, auditory noise (NL σ_{Aud} 0.12, HL σ_{Aud} 0.12) and visual noise (NL σ_{Vis} 0.48, HL σ_{Vis} 0.47) remained stable across load. For the visual distractor group, auditory noise remained stable (NL σ_{Aud} 0.15, HL σ_{Aud} 0.17); however, visual noise increased (NL σ_{Vis} 0.52, HL σ_{Vis} 0.67). An

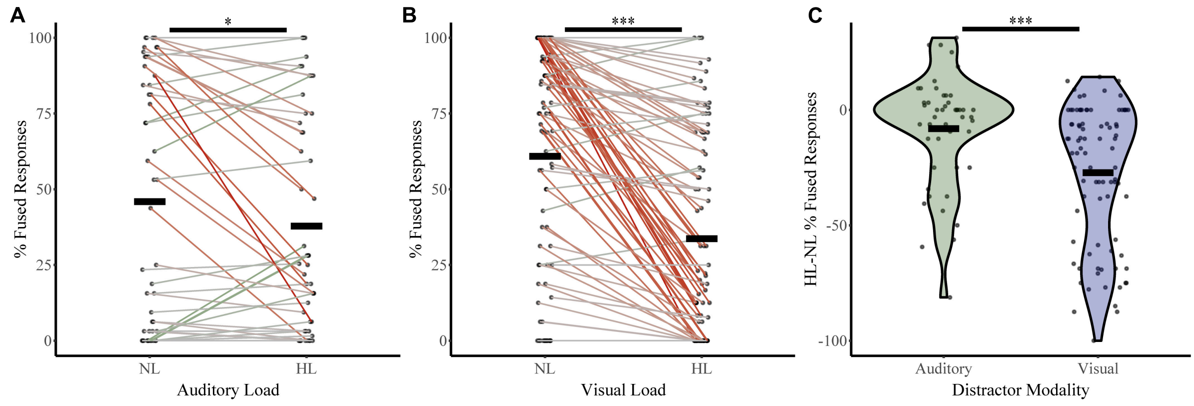


FIGURE 3
McGurk fused reports for no load (NL) and high load (HL) bocks. The percent of fused reports ("da" or "tha") during NL and HL blocks are shown for each participant for the auditory distractor **(A)** and visual distractor **(B)** tasks. Horizontal black bars indicate group averages. Colored lines connect individual percent fused reports across each block with a green line indicating an increase in fused reports from NL to HL and a red line indicating a decrease. The difference in percent fused reports across load for rapid serial visual presentation (RSVP) and rapid serial auditory presentation (RSAP) tasks is shown in panel **(C)**. ***Indicates $p < 0.001$ and *indicates $p < 0.05$.

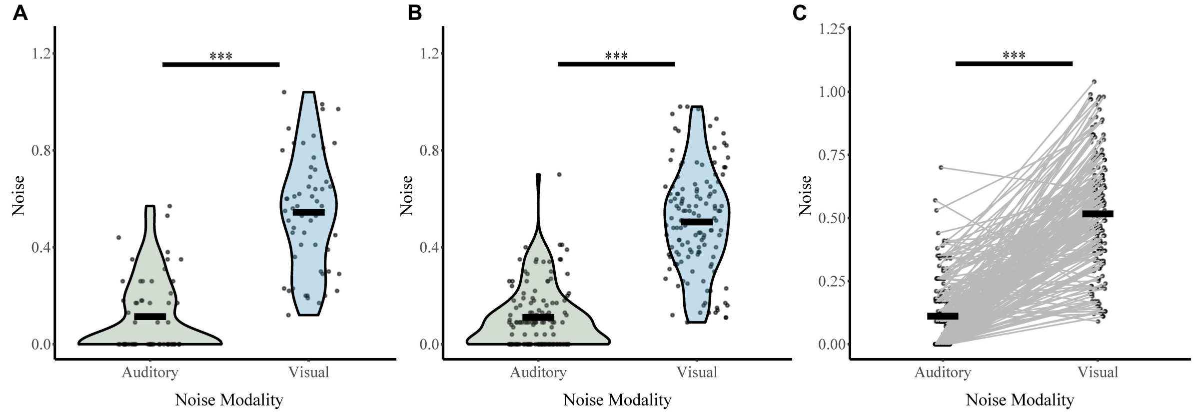


FIGURE 4
Sensory noise for distractor free blocks. Auditory and visual sensory noise is shown separately for the auditory distractor **(A)** and visual distractor **(B)** groups. Horizontal black bars indicate group averages, and violin plots display the distribution of sensory noise in each modality for each task. Panel **(C)** shows auditory and visual sensory noise for all participants connected for each participant with straight lines. ***Indicates $p < 0.001$.

RMANOVA of sensory noise with noise modality and load (NL or HL) as within-subjects factors and distractor modality as a between-subjects factor revealed significant main effects of noise modality ($F_{1,133} = 414.836$, $p = 1.03 \times 10^{-42}$, partial $\eta^2 = 0.757$), load ($F_{1,133} = 5.702$, $p = 0.02$, partial $\eta^2 = 0.041$), and distractor modality ($F_{\ 1,133} = 11.816$, $p = 0.001$, partial $\eta^2 = 0.082$). There were also significant interactions between load and distractor modality ($F_{1,133} = 8.06$, $p = 0.005$, partial $\eta^2 = 0.057$) and a three-way interaction between noise modality, load, and distractor modality ($F_{1,133} = 7.612$, $p = 0.007$, partial $\eta^2 = 0.054$). The interaction between distractor modality and noise modality approached significance ($F_{1,133} = 3.890$, $p = 0.051$, partial $\eta^2 = 0.028$). *Post-hoc* analyses using t-tests and non-parametric assessments corroborated these findings. Visual noise increased from no load to high load in visual modality only ($t_{86} = -4.78$, $p = 7.28 \times 10^{-6}$, Cohen's d: -0.51; Bootstrapped 95% CI: -0.22 to -0.09, $p = 2.00 \times 10^{-4}$; $W = 2{,}928$, $z = 4.29$, $p = 1.77 \times 10^{-5}$; $BF = 0.01$). However, visual noise did not significantly change from no load to high load with auditory distractors ($t_{47} = 0.53$, $p = 0.60$, Cohen's d: 0.08; Bootstrapped 95% CI: -0.04 to 0.07, $p = 0.60$; $W = 541.5$, $z = -0.238$, $p = 0.81$; $BF = 7.71$). As follows, change in visual noise was higher with visual distractors than auditory distractors ($t_{131} = 4.0$, $p = 1.04 \times 10^{-4}$, Cohen's $d = 0.63$, Bootstrapped 95% CI: 0.09–0.26; $p = 2.00 \times 10^{-4}$; $\mathrm{U}_{N,AudDist:48;N,VisDist:87} = 1{,}329$, $z = -3.49$, $p = 4.85 \times 10^{-4}$; Bayesian $t_{133} = 3.52$, $p = 1.0 \times 10^{-3}$, $BF = 0.025$). Further, auditory noise did not significantly change from no load to high

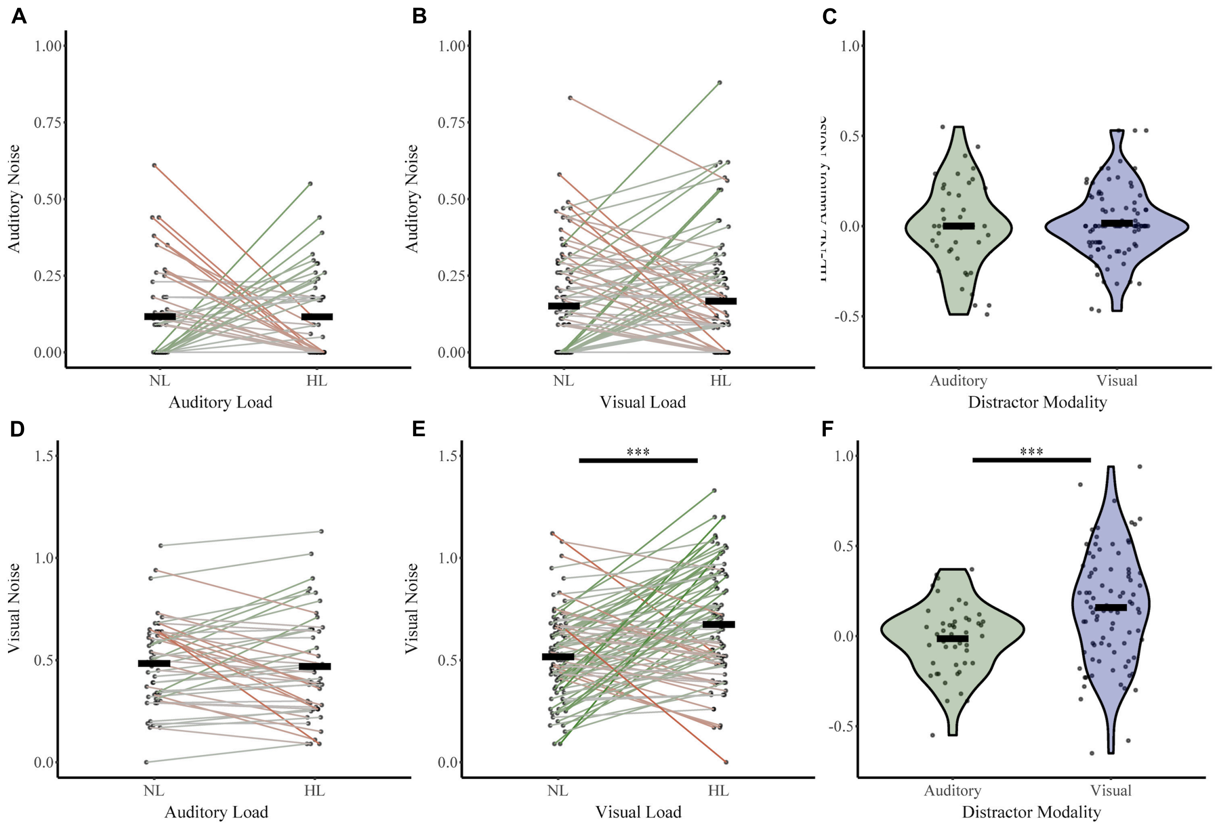


FIGURE 5

Changes in sensory noise across perceptual load. Auditory noise does not change with increasing auditory **(A)** or visual **(B)** perceptual load. Visual noise increases with increasing visual **(E)** but not auditory **(D)** noise. HL-NL differences in auditory noise **(C)** and visual noise **(F)** confirm that visual load selectively increases visual noise. Horizontal black bars indicate group averages, and violin plots display the distribution of HL-NL sensory noise differences for each distractor and noise modality. ***Indicates $p < 0.001$.

load with either visual ($t_{86} = -0.74$, $p = 0.46$, Cohen's d: -0.08; Bootstrapped 95% CI: -0.06 to 0.03, $p = 0.47$; $W = 1{,}236$, $z = 0.606$, $p = 0.545$; $BF = 9.05$) or auditory distractors ($t_{47} = 0.01$, $p = 1.00$, Cohen's d: 1.0×10^{-3}; Bootstrapped 95% CI: -0.07 to 0.07, $p = 1.00$; $W = 331.500$, $z = -0.024$, $p = 0.98$; $BF = 8.86$). The difference in auditory noise from no load to high load did not significantly differ between distractor modality ($t_{84} = 0.40$, $p = 0.69$, Cohen's $d = 0.08$; Bootstrapped 95% CI: -0.06 to 0.09, $p = 0.70$; $U_{N,AudDist:48;N,VisDist:87} = 2{,}025$, $z = -0.29$, $p = 0.77$; $t_{133} = 0.42$, $p = 0.68$, $BF = 6.17$) Collectively, these findings indicate that attentional increases in sensory noise are specific to visual noise with increasing visual load only.

3.3. Multiple linear regressions analysis models

3.3.1. Baseline McGurk reports

We constructed a multiple linear regression model to determine which sensory noise measures (auditory noise, visual noise, or a combination of both) best predicted baseline McGurk reports. Distractor Modality was included in the model because our RMANOVA analyses (described above) identified it as a significant factor. While neither visual noise ($r_{134} = 0.028$, $p = 0.701$) nor auditory noise ($r_{134} = 0.118$, $p = 0.104$) correlated with baseline McGurk reports, combined noise did significantly correlate with baseline McGurk reports ($r_{134} = -0.172$, $p = 0.017$). Thus, we constructed a multiple regression model to predict baseline McGurk reports with distractor modality and combined noise as factors (Table 1). A significant relationship was found ($F_{2,189} = 13.24$, $p = 4.16 \times 10^{-6}$) with an R^2 of 0.123. Baseline McGurk reports were significantly predicted by distractor modality ($\beta = -0.306$, $t = -4.49$, $p = 1.26 \times 10^{-5}$; bootstrap $p = 0.0002$) and combined noise ($\beta = -0.150$, $t = -2.20$, $p = 0.029$; bootstrap $p = 0.049$; Figure 6A). Neither auditory noise ($\Delta F_{1,188} = 0.05$, $p = 0.817$; $\Delta R^2 = 0.0002$) nor visual noise ($\Delta F_{1,187} = 3.25$, $p = 0.073$; $\Delta R^2 = 0.015$) significantly increased the predictability of this multiple regression model when added in stepwise fashion, confirming the relative importance of combined noise in predicting baseline McGurk perceptions.

TABLE 1 Multiple linear regression: Baseline McGurk reports showing predictive power of distractor modality and combined noise on baseline McGurk perception.

Predictor	Unstandardized coefficients	SE	95% CI (Bootstrapped)		β	*p*-value
			Lower	Upper		
Intercept	79.16	5.96	66.93	91.42		0.0002
Distractor modality	−26.14	5.83	−37.64	−14.12	−0.306	0.0002
Combined noise	−16.12	7.57	−33.26	−0.40	−0.150	0.049

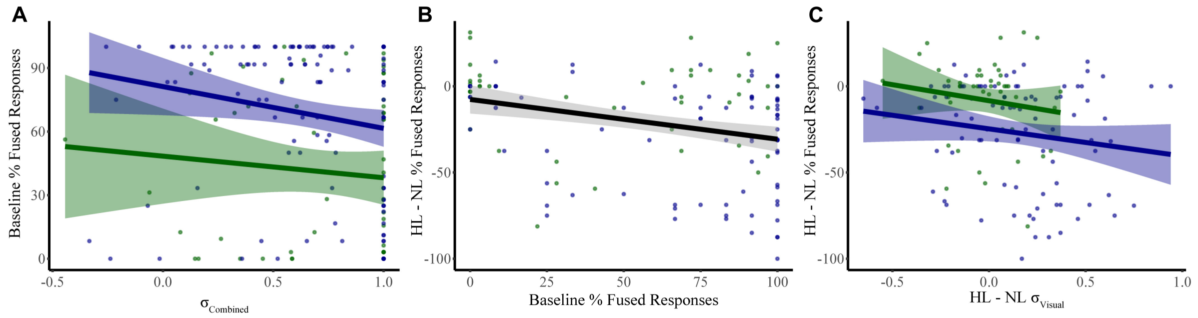


FIGURE 6
Significant predictors of McGurk fused reports. Our first model identified combined sensory noise and distractor modality as significant predictors of fused reports during baseline conditions (distractor free) **(A)**. Changes in fused reports from NL to HL conditions were related to both baseline McGurk fused reports **(B)** and to the change in visual noise **(C)**. Shaded regions reflect the 95% confidence interval for the regression.

3.3.2. Dual task McGurk reports

We constructed a multiple linear regression model to determine which factors contributed to changes in McGurk reports with increasing perceptual load. To determine which factors to include, we performed separate RMANOVAs with visual noise, auditory noise, or combined noise as dependent variables, perceptual load as a within-subjects factor, and distractor modality as a between-subjects factor. For visual noise, there was a significant main effect of load ($F_{1,133} = 8.51$, $p = 0.004$, partial $\eta^2 = 0.060$) and distractor modality ($F_{1,133} = 11.079$, $p = 0.001$, partial $\eta^2 = 0.077$) as well as a significant interaction between load and distractor modality ($F_{1,133} = 12.38$, $p = 0.001$, partial $\eta^2 = 0.085$). There were no significant effects for auditory noise (load: $F_{1,133} = 0.164$, $p = 0.686$, partial $\eta^2 = 0.001$; distractor modality: $F_{1,133} = 3.064$, $p = 0.082$, partial $\eta^2 = 0.023$; interaction: $F_{1,133} = 0.173$, $p = 0.678$, partial $\eta^2 = 0.001$) or combined noise (load: $F_{1,133} = 0.720$, $p = 0.398$, partial $\eta^2 = 0.005$; distractor modality: $F_{1,133} = 0.421$, $p = 0.517$, partial $\eta^2 = 0.003$; interaction: $F_{1,133} = 0.101$, $p = 0.751$, partial $\eta^2 = 0.001$). Additionally, the change in McGurk reports from no load to high load significantly correlated with the change in Visual Noise from no load to high load ($r_{134} = -0.235$, $p = 0.006$) and not change in Auditory ($r_{134} = -0.085$, $p = 0.330$) or change in Combined Noise ($r_{134} = -0.044$, $p = 0.615$). Collectively, these results suggest that changes in visual noise across load best explain changes in McGurk perception with increasing load as compared to other measures of sensory noise. Thus, we constructed a multiple linear regression model with change in McGurk reports from no load to high load as the dependent variable and the following potential explanatory variables: baseline McGurk reports, change in visual noise, and distractor modality (**Table 2**). A significant relationship was found ($F_{3,131} = 10.32$, $p = 3.81 \times 10^{-6}$) with an R^2 of 0.191. Change in McGurk reports was significantly predicted by baseline McGurk reports ($\beta = -0.276$, $t = -3.42$, $p = 0.001$; bootstrap $p = 4.00 \times 10^{-4}$; **Figure 6B**), Distractor Modality ($\beta = 0.197$, $t = 2.33$, $p = 0.021$; bootstrap $p = 0.008$), change in Visual Noise ($\beta = -0.184$, $t = -2.24$, $p = 0.027$; bootstrap $p = 0.022$; **Figure 6C**). Neither change in auditory noise ($\Delta F_{1,130} = 0.20$, $p = 0.654$; $\Delta R^2 = 0.001$) nor change in combined noise ($\Delta F_{1,129} = 0.18$, $p = 0.672$; $\Delta R^2 = 0.001$) increased the predictability of this multiple regression model when added in stepwise fashion, confirming the relative importance of changes in visual noise predicting attentional disruptions to McGurk perceptions.

4. Discussion

The present study investigated whether variations in sensory noise could explain the impact of attention on multisensory integration of speech stimuli and to what extent this mechanism operates in a modality-specific manner. To

TABLE 2 Multiple linear regression: Dual-task McGurk reports showing predictive power of distractor modality, baseline McGurk, and ∆visual noise on ∆McGurk responses from NL to HL.

Predictor	Unstandardized coefficients	SE	95% CI (Bootstrap)		β	*p*-value (Bootstrap)
			Lower	Upper		
Intercept	−11.61	4.95	−19.31	−5.07		0.001
Baseline McGurk	−0.20	0.06	−0.30	−0.10	−0.276	0.0004
Distractor modality	12.10	5.20	3.00	20.91	0.197	0.008
∆Visual noise	−19.13	8.60	−36.43	−3.90	−0.184	0.022

examine within-modality effects, we created a novel method of measuring sensory noise based on response variability in unisensory trials. Importantly, this method expands on previous models, allowing us to investigate the effects of visual and auditory noise independently from one another. Consistent with other computational models of multisensory speech integration, the overwhelming majority of participants had higher visual noise compared to auditory (Massaro, 1999; Ma et al., 2009; Magnotti and Beauchamp, 2015, 2017; Magnotti et al., 2020). Additionally, our combined sensory noise measure, which is the direct equivalent of the sensory noise ratio in the CIMS model (Magnotti and Beauchamp, 2017; Magnotti et al., 2020), was a better predictor of baseline McGurk reports than visual or auditory noise alone. These findings are strongly aligned with other computational measures of sensory noise and lend evidence to the overall importance of sensory noise for multisensory integration. The novel method of estimating sensory noise separately for each modality provides additional functionality to current models of multisensory speech integration which primarily rely on the relative levels of visual and auditory noise but do not permit either to vary independently (Magnotti and Beauchamp, 2017). These within-modality measures of sensory noise allowed us to identify that changes in visual noise, specifically, were associated with attentional modulations to multisensory speech perception. Increases in visual load led to increased visual noise and decreased McGurk perception. Correspondingly, changes in visual noise were predictive of changes to McGurk reports across load. These findings suggest that attention alters the encoding of visual speech information and that attention may impact sensory noise in a modality-specific manner. Unfortunately, our method of calculating sensory noise resulted in many participants having an auditory noise value of zero even under high perceptual load, suggesting that this method may not be sensitive enough to estimate very low levels of sensory noise. However, it can accurately determine the individual contributions of and changes to auditory and visual noise on multisensory integration.

Our results strongly indicate that modulations of attention differentially impact multisensory speech perception depending on the sensory modality of the attentional manipulation. While we found striking increases in visual noise with increasing visual load, we did not find corresponding increases in auditory noise with increasing auditory load suggesting a separate mechanism by which auditory attention influences multisensory speech integration. Additionally, while increasing perceptual load led to decreased McGurk reports for both visual and auditory secondary tasks, this effect was more pronounced for the visual task suggesting that alterations to visual attention may have a heightened impact on multisensory speech integration. Because the auditory and visual secondary tasks differed in ways other than their modality, we cannot eliminate the possibility that these differences account for our observed modality effects. We hypothesize that our visual secondary task engages featural attention, and although our secondary auditory task asked participants to identify auditory features (i.e., pitch and duration), we suspect that participants listened for melodic or rhythmic indicators of targets which may have engaged object-based attention. Future research is needed to investigate the relative contributions of distractor modality and type of attentional manipulation on multisensory speech integration. Another potential explanation for distractor modality effects is differential patterns of eye movements. Gaze behavior has been shown to influence the McGurk effect (Paré et al., 2003; Gurler et al., 2015; Jensen et al., 2018; Wahn et al., 2021). Because eye movements were not monitored during this study, future research is needed to investigate whether gaze behavior may explain modality differences in the impact of the secondary task on multisensory speech integration. Surprisingly, McGurk reports differed in the distractor-free condition across auditory vs. visual secondary task groups even though the tasks were identical. This suggests that the sensory modality of a secondary task may influence multisensory speech perception even when not concurrently presented. Approximately 70% of participants completed the distractor-free block after a low load or high load block, suggesting that our secondary task may prime attention to its corresponding modality and subsequently alter speech integration. Interestingly, we did not find differences in sensory noise across distractor modality in the distractor-free condition. This implies that any task context effects may lead to changes in participants' priors or relative weighing of auditory vs. visual speech information (Shams et al., 2005; Kayser and Shams, 2015;

Magnotti and Beauchamp, 2017; Magnotti et al., 2020). The current study was not designed to assess order effects; thus, future research is needed to fully investigate modality-specific priming effects and to elucidate the mechanisms by which they may influence multisensory speech perception.

The results of this study inform our understanding of the mechanisms by which attention influences multisensory processing. Multisensory speech integration relies on both extensive processing of the auditory and visual speech signal and convergence of auditory and visual pathways onto multisensory cortical sites such as the Superior Temporal Sulcus (STS) (Beauchamp et al., 2004, 2010; Callan et al., 2004; Nath and Beauchamp, 2011, 2012; O'Sullivan et al., 2019, 2021; Ahmed et al., 2021; Nidiffer et al., 2021). Additionally, the functional connectivity between STS and unisensory cortices differs according to the reliability of the corresponding unisensory information (e.g., increased visual reliability will lead to increased functional connectivity between visual cortex and STS) (Nath and Beauchamp, 2011). Our findings suggest that increasing visual load leads to disrupted encoding of the visual speech signal which then leads to a deweighting of visual information potentially through decreased functional connectivity between the STS and visual cortex. Interestingly, increasing auditory load does not appear to disrupt multisensory speech integration through the same mechanism. Ahmed et al. found that attention favors integration at later stages of speech processing (Ahmed et al., 2021) suggesting that our secondary auditory task may disrupt later stages of integrative processing. Future research utilizing neuroimaging methodology is needed to link behavioral estimates of sensory noise to specific neural mechanisms.

Identifying the specific neural mechanisms by which top-down cognitive factors shape multisensory processing is important for our understanding of how multisensory integration functions in realistic contexts and across individual differences. For example, older adults exhibit either intact, enhanced, or shifted patterns of multisensory integration depending on the task utilized in the study (Hugenschmidt et al., 2009; Freiherr et al., 2013; de Dieuleveult et al., 2017; Parker and Robinson, 2018). Interestingly, several studies have shown altered sensory dominance and weighting of unisensory information in older adults when compared to younger adults (Murray et al., 2018; Jones and Noppeney, 2021). Within-modality measures of sensory noise as described in this study may help to illuminate the reasons why certain multisensory stimuli and tasks lead to differences in the multisensory effects observed in the aging population. Cognitive control mechanisms are also known to decline with healthy aging, and manipulations of attention (e.g., dual-task designs) consistently have a larger impact on the elderly (Mahoney et al., 2012; Carr et al., 2019; Ward et al., 2021). Currently, there is a gap in knowledge on how attention may alter relative sensory weighting in older adults that could be addressed by utilizing the experimental design described in this study. Addressing this gap in knowledge could improve our understanding of multisensory speech integration in normal aging and with sensory loss (Peter et al., 2019; Dias et al., 2021) as well as current multisensory screening tools for assessing risks for falls in the elderly (Mahoney et al., 2019; Zhang et al., 2020). In addition to healthy aging, many developmental disorders are characterized by disruptions to both multisensory functioning and attention, and these neurological processes may interact to worsen the severity of these disorders (Belmonte and Yurgelun-Todd, 2003; de Jong et al., 2010; Kwakye et al., 2011; Magnée et al., 2011; Harrar et al., 2014; Krause, 2015; Mayer et al., 2015; Noel et al., 2018b). Previous research indicates that deficits in processing both speech (van Laarhoven et al., 2019) and non-speech (Leekam et al., 2007) stimuli were present in subjects on the autism spectrum. Sensory noise and its interactions with attention may contribute to differences in ASD sensory processing beyond stimulus signal-to-noise ratio or general neural noise. Investigating these mechanisms may help us understand and identify disruptions in the relationship between multisensory integration and attention, inspiring new strategies for interventions to address altered functioning in these disorders.

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 Oberlin College Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

Author contributions

LK, VF, CD, EP, WK, and CN contributed to the conception and design of the project and wrote the manuscript. LK and CD collected the data for the project. LK, VF, and CN developed the sensory noise analysis, and analyzed and interpreted the final data for the manuscript. CD, EP, and WK contributed to initial data analysis and interpretation. All authors contributed to the article and approved the submitted version.

Acknowledgments

We would like to acknowledge the contributions of Olivia Salas, Kyla Gibney, Brady Eggleston, and Enimielen

Aligbe to some of the data collection for this study. We would also like to thank the Oberlin College Research Fellowship and the Oberlin College Office of Foundation, Government, and Corporate Grants for their support of this study.

Conflict of interest

CD is a current employee of F. Hoffman-La Roche Ltd. and Genentech, Inc. CD's contributions to this study were completed while she attended Oberlin College.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2022.1027335/full#supplementary-material

SUPPLEMENTARY FIGURE 1

Performance on low-load (LL) and high-load (HL) distractor tasks. Participants scored significantly higher during the LL conditions than the HL conditions for both auditory (**A**; $LL = 92.6$; $HL = 78.6$; $t_{44} = 10.39$, $p = 2.04 \times 10^{-13}$, Cohen's $d = 1.79$) and visual (**B**; $LL = 94.0$; $HL = 85.4$; $t_{76} = 13.81$, $p = 2.02 \times 10^{-22}$, Cohen's $d = 1.48$) distractor tasks.

SUPPLEMENTARY FIGURE 2

Performance on unisensory and congruent multisensory trials for no load (NL), low load (LL), and high load (HL) blocks. Percent correct syllable identification for visual-only, auditory-only, and multisensory congruent trials for the auditory distractor **(A)** and visual distractor **(B)** tasks. An RMANOVA revealed that distractor modality ($F_{1,128} = 12.8$, $p = 4.91 \times 10^{-4}$, partial $\eta^2 = 0.091$), perceptual load ($F_{2,256} = 6.7$, $p = 0.001$, partial $\eta^2 = 0.050$), and syllable modality ($F_{2,256} = 1040.5$, $p = 1.16 \times 10^{-123}$, partial $\eta^2 = 0.890$) significantly altered accuracy.

SUPPLEMENTARY FIGURE 3

McGurk fused reports for no load (NL), low load (LL), and high load (HL). The percent of fused reports ("da" or "tha") during each block are shown for auditory distractor **(A)** and visual distractor **(B)** tasks. Horizontal bars indicate group averages. Colored lines connect individual percent fused reports across each block. Green lines indicate increased in fused reports and a red line indicates a decrease in fused reports. An RMANOVA revealed that both perceptual load ($F_{2,256} = 22.5$, $p = 9.90 \times 10^{-10}$, partial $\eta^2 = 0.148$) and the interaction between load and distractor modality ($F_{2,256} = 4.7$, $p = 0.010$, partial $\eta^2 = 0.035$) significantly altered percent McGurk reports.

References

Ahmed, F., Nidiffer, A. R., O'Sullivan, A. E., Zuk, N. J., and Lalor, E. C. (2021). The integration of continuous audio and visual speech in a cocktail-party environment depends on attention. *bioRxiv* [Preprint] doi: 10.1101/2021.02.10.430634

Al-Aidroos, N., Said, C. P., and Turk-Browne, N. B. (2012). Top-down attention switches coupling between low-level and high-level areas of human visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 109, 14675–14680. doi: 10.1073/pnas.1202095109

Alsius, A., Möttönen, R., Sams, M. E., Soto-Faraco, S., and Tiippana, K. (2014). Effect of attentional load on audiovisual speech perception: Evidence from ERPs. *Front. Psychol.* 5:727. doi: 10.3389/fpsyg.2014.00727

Alsius, A., Navarra, J., Campbell, R., and Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Curr. Biol.* 15, 839–843. doi: 10.1016/j.cub.2005.03.046

Alsius, A., Navarra, J., and Soto-Faraco, S. (2007). Attention to touch weakens audiovisual speech integration. *Exp. Brain Res.* 183, 399–404. doi: 10.1007/s00221-007-1110-1

Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., and Martin, A. (2004). Unraveling multisensory integration: Patchy organization within human STS multisensory cortex. *Nat. Neurosci.* 7, 1190–1192. doi: 10.1038/nn1333

Beauchamp, M. S., Nath, A. R., and Pasalar, S. (2010). FMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *J. Neurosci.* 30, 2414–2417. doi: 10.1523/JNEUROSCI.4865-09.2010

Belmonte, M. K., and Yurgelun-Todd, D. A. (2003). Functional anatomy of impaired selective attention and compensatory processing in autism. *Brain Res. Cogn. Brain Res.* 17, 651–664. doi: 10.1016/S0926-6410(03)00189-7

Bobrowski, O., Meir, R., and Eldar, Y. C. (2009). Bayesian filtering in spiking neural networks: Noise, adaptation, and multisensory integration. *Neural. Comput.* 21, 1277–1320. doi: 10.1162/neco.2008.01-08-692

Bonato, M., Spironelli, C., Lisi, M., Priftis, K., and Zorzi, M. (2015). Effects of multimodal load on spatial monitoring as revealed by ERPs. *PLoS One* 10:e0136719. doi: 10.1371/journal.pone.0136719

Burns, J. K., and Blohm, G. (2010). Multi-sensory weights depend on contextual noise in reference frame transformations. *Front. Hum. Neurosci.* 4:221. doi: 10.3389/fnhum.2010.00221

Callan, D. E., Jones, J. A., Munhall, K., Kroos, C., Callan, A. M., and Vatikiotis-Bateson, E. (2004). Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. *J. Cogn. Neurosci.* 16, 805–816. doi: 10.1162/089892904970771

Carr, S., Pichora-Fuller, M. K., Li, K. Z. H., Phillips, N., and Campos, J. L. (2019). Multisensory, multi-tasking performance of older adults with and without subjective cognitive decline. *Multisens. Res.* 32, 797–829. doi: 10.1163/22134808-20191426

Cascio, C. N., O'Donnell, M. B., Tinney, F. J., Lieberman, M. D., Taylor, S. E., Strecher, V. J., et al. (2016). Self-affirmation activates brain systems associated with self-related processing and reward and is reinforced by future orientation. *Soc. Cogn. Affect. Neurosci.* 11, 621–629. doi: 10.1093/scan/nsv136

Dean, C. L., Eggleston, B. A., Gibney, K. D., Aligbe, E., Blackwell, M., and Kwakye, L. D. (2017). Auditory and visual distractors disrupt multisensory temporal acuity in the crossmodal temporal order judgment task. *PLoS One* 12:e0179564. doi: 10.1371/journal.pone.0179564

Deneve, S., and Pouget, A. (2004). Bayesian multisensory integration and cross-modal spatial links. *J. Physiol. Paris* 98, 249–258. doi: 10.1016/j.jphysparis.2004.03.011

Dias, J. W., McClaskey, C. M., and Harris, K. C. (2021). Audiovisual speech is more than the sum of its parts: Auditory-visual superadditivity compensates for age-related declines in audible and lipread speech intelligibility. *Psychol. Aging* 36, 520–530. doi: 10.1037/pag0000613 de Dieuleveult, A. L., Siemonsma, P. C., van Erp, J. B. F., and Brouwer, A.-M. (2017). Effects of aging in multisensory integration: A systematic review. *Front. Aging Neurosci.* 9:80. doi: 10.3389/fnagi.2017.00080

Driver, J. (2001). A selective review of selective attention research from the past century. *Br. J. Psychol.* 1, 53–78. doi: 10.1348/000712601162103

de Jong, J. J., Hodiamont, P. P. G., and de Gelder, B. (2010). Modality-specific attention and multisensory integration of emotions in schizophrenia: Reduced regulatory effects. *Schizophr. Res.* 122, 136–143. doi: 10.1016/j.schres.2010.04.010

Engel, A. K., Senkowski, D., and Schneider, T. R. (2012). "Multisensory integration through neural coherence," in *The Neural Bases of Multisensory Processes* Frontiers in Neuroscience, eds M. M. Murray, M. T. Wallace, M. M. Murray, M. T. Wallace, M. M. Murray, M. T. Wallace, et al. (Boca Raton, FL: CRC Press).

Ernst, M. O., and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433. doi: 10.1038/415429a

Freiherr, J., Lundström, J. N., Habel, U., and Reetz, K. (2013). Multisensory integration mechanisms during aging. *Front. Hum. Neurosci.* 7:863. doi: 10.3389/fnhum.2013.00863

Friese, U., Daume, J., Göschl, F., König, P., Wang, P., and Engel, A. K. (2016). Oscillatory brain activity during multisensory attention reflects activation, disinhibition, and cognitive control. *Sci. Rep.* 6:32775. doi: 10.1038/srep32775

Gibney, K. D., Aligbe, E., Eggleston, B. A., Nunes, S. R., Kerkhoff, W. G., Dean, C. L., et al. (2017). Visual distractors disrupt audiovisual integration regardless of stimulus complexity. *Front. Integr. Neurosci.* 11:1. doi: 10.3389/fnint.2017.00001

Gomez-Ramirez, M., Kelly, S. P., Molholm, S., Sehatpour, P., Schwartz, T. H., and Foxe, J. J. (2011). Oscillatory sensory selection mechanisms during intersensory attention to rhythmic auditory and visual inputs: A human electrocorticographic investigation. *J. Neurosci.* 31, 18556–18567. doi: 10.1523/JNEUROSCI.2164-11.2011

Gurler, D., Doyle, N., Walker, E., Magnotti, J., and Beauchamp, M. (2015). A link between individual differences in multisensory speech perception and eye movements. *Atten. Percept. Psychophys.* 77, 1333–1341. doi: 10.3758/s13414-014-0821-1

Harrar, V., Tammam, J., Pérez-Bellido, A., Pitt, A., Stein, J., and Spence, C. (2014). Multisensory integration and attention in developmental dyslexia. *Curr. Biol.* 24, 531–535. doi: 10.1016/j.cub.2014.01.029

Hugenschmidt, C. E., Mozolic, J. L., and Laurienti, P. J. (2009). Suppression of multisensory integration by modality-specific attention in aging. *Neuroreport* 20, 349–353. doi: 10.1097/WNR.0b013e328323ab07

Jensen, A., Merz, S., Spence, C., and Frings, C. (2018). Overt spatial attention modulates multisensory selection. *J. Exp. Psychol. Hum. Percept. Perform.* 45, 174–188. doi: 10.1037/xhp0000595

Jones, S. A., and Noppeney, U. (2021). Ageing and multisensory integration: A review of the evidence, and a computational perspective. *Cortex* 138, 1–23. doi: 10.1016/j.cortex.2021.02.001

Kayser, C., and Shams, L. (2015). Multisensory causal inference in the brain. *PLoS Biol.* 13:e1002075. doi: 10.1371/journal.pbio.1002075

Keil, J., Pomper, U., and Senkowski, D. (2016). Distinct patterns of local oscillatory activity and functional connectivity underlie intersensory attention and temporal prediction. *Cortex* 74, 277–288. doi: 10.1016/j.cortex.2015.10.023

Koelewijn, T., Bronkhorst, A., and Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta Psychol.* 134, 372–384. doi: 10.1016/j.actpsy.2010.03.010

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., and Shams, L. (2007). Causal inference in multisensory perception. *PLoS One* 2:e943. doi: 10.1371/journal.pone.0000943

Krause, M. B. (2015). Pay Attention!: Sluggish multisensory attentional shifting as a core deficit in developmental dyslexia. *Dyslexia* 21, 285–303. doi: 10.1002/dys.1505

Kwakye, L. D., Foss-Feig, J. H., Cascio, C. J., Stone, W. L., and Wallace, M. T. (2011). Altered auditory and multisensory temporal processing in autism spectrum disorders. *Front. Integr. Neurosci.* 4:129. doi: 10.3389/fnint.2010.00129

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320, 110–113. doi: 10.1126/science.1154735

Lalonde, K., and Werner, L. A. (2019). Perception of incongruent audiovisual English consonants. *PLoS One* 14:e0213588. doi: 10.1371/journal.pone.0213588

Lavie, N., Ro, T., and Russell, C. (2003). The role of perceptual load in processing distractor faces. *Psychol. Sci.* 14, 510–515. doi: 10.1111/1467-9280.03453

Leekam, S. R., Nieto, C., Libby, S. J., Wing, L., and Gould, J. (2007). Describing the sensory abnormalities of children and adults with autism. *J. Autism Dev. Disord.* 37, 894–910. doi: 10.1007/s10803-006-0218-7

Ling, S., Liu, T., and Carrasco, M. (2009). How spatial and feature-based attention affect the gain and tuning of population responses. *Vision Res.* 49, 1194–1204. doi: 10.1016/j.visres.2008.05.025

Magnée, M. J. C. M., de Gelder, B., van Engeland, H., and Kemner, C. (2011). Multisensory integration and attention in autism spectrum disorder: Evidence from event-related potentials. *PLoS One* 6:e24196. doi: 10.1371/journal.pone.0024196

Magnotti, J. F., and Beauchamp, M. S. (2017). A causal inference model explains perception of the mcgurk effect and other incongruent audiovisual speech. *PLoS Comput. Biol.* 13:e1005229. doi: 10.1371/journal.pcbi.1005229

Magnotti, J. F., and Beauchamp, M. S. (2015). The noisy encoding of disparity model of the McGurk effect. *Psychon. Bull. Rev.* 22, 701–709. doi: 10.3758/s13423-014-0722-2

Magnotti, J. F., Dzeda, K. B., Wegner-Clemens, K., Rennig, J., and Beauchamp, M. S. (2020). Weak observer-level correlation and strong stimulus-level correlation between the McGurk effect and audiovisual speech-in-noise: A causal inference explanation. *Cortex* 133, 371–383. doi: 10.1016/j.cortex.2020.10.002

Magnotti, J. F., Ma, W. J., and Beauchamp, M. S. (2013). Causal inference of asynchronous audiovisual speech. *Front. Psychol.* 4:798. doi: 10.3389/fpsyg.2013.00798

Magnotti, J. F., Smith, K. B., Salinas, M., Mays, J., Zhu, L. L., and Beauchamp, M. S. (2018). A causal inference explanation for enhancement of multisensory integration by co-articulation. *Sci. Rep.* 8:18032. doi: 10.1038/s41598-018-36772-8

Mahoney, J. R., Cotton, K., and Verghese, J. (2019). Multisensory integration predicts balance and falls in older adults. *J. Gerontol. A Biol. Sci. Med. Sci.* 74, 1429–1435. doi: 10.1093/gerona/gly245

Mahoney, J. R., Verghese, J., Dumas, K., Wang, C., and Holtzer, R. (2012). The effect of multisensory cues on attention in aging. *Brain Res.* 1472, 63–73. doi: 10.1016/j.brainres.2012.07.014

Mangun, G. R. (1995). Neural mechanisms of visual selective attention. *Psychophysiology* 32, 4–18. doi: 10.1111/j.1469-8986.1995.tb03400.x

Massaro, D. W. (1999). Speechreading: Illusion or window into pattern recognition. *Trends Cogn. Sci.* 3, 310–317. doi: 10.1016/S1364-6613(99)01360-1

Mayer, A. R., Hanlon, F. M., Teshiba, T. M., Klimaj, S. D., Ling, J. M., Dodd, A. B., et al. (2015). An fMRI study of multimodal selective attention in schizophrenia. *Br. J. Psychiatry* 207, 420–428. doi: 10.1192/bjp.bp.114.155499

Ma, W. J., Zhou, X., Ross, L. A., Foxe, J. J., and Parra, L. C. (2009). Lip-reading aids word recognition most in moderate noise: A bayesian explanation using high-dimensional feature space. *PLoS One* 4:e4638. doi: 10.1371/journal.pone.0004638

Mcgurk, H., and Macdonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0

Mishra, J., and Gazzaley, A. (2012). Attention distributed across sensory modalities enhances perceptual performance. *J. Neurosci.* 32, 12294–12302. doi: 10.1523/JNEUROSCI.0867-12.2012

Mishra, J., Martínez, A., and Hillyard, S. A. (2010). Effect of attention on early cortical processes associated with the sound-induced extra flash illusion. *J. Cogn. Neurosci.* 22, 1714–1729. doi: 10.1162/jocn.2009.21295

Mitchell, J. F., Sundberg, K. A., and Reynolds, J. H. (2007). Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron* 55, 131–141. doi: 10.1016/j.neuron.2007.06.018

Morís Fernández, L., Visser, M., Ventura-Campos, N., Ávila, C., and Soto-Faraco, S. (2015). Top-down attention regulates the neural expression of audiovisual integration. *Neuroimage* 119, 272–285. doi: 10.1016/j.neuroimage.2015.06.052

Mozolic, J. L., Hugenschmidt, C. E., Peiffer, A. M., and Laurienti, P. J. (2008). Modality-specific selective attention attenuates multisensory integration. *Exp. Brain Res.* 184, 39–52. doi: 10.1007/s00221-007-1080-3

Murray, M. M., Eardley, A. F., Edginton, T., Oyekan, R., Smyth, E., and Matusz, P. J. (2018). Sensory dominance and multisensory integration as screening tools in aging. *Sci. Rep.* 8:8901. doi: 10.1038/s41598-018-27288-2

Nath, A. R., and Beauchamp, M. S. (2012). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *Neuroimage* 59, 781–787. doi: 10.1016/j.neuroimage.2011.07.024

Nath, A. R., and Beauchamp, M. S. (2011). Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. *J. Neurosci.* 31, 1704–1714. doi: 10.1523/JNEUROSCI.4853-10.2011

Nidiffer, A. R., Cao, C. Z., O'Sullivan, A. E., and Lalor, E. C. (2021). A linguistic representation in the visual system underlies successful lipreading. *bioRxiv* [Preprint] doi: 10.1101/2021.02.09.430299

Noel, J.-P., Serino, A., and Wallace, M. T. (2018a). Increased neural strength and reliability to audiovisual stimuli at the boundary of peripersonal space. *J. Cogn. Neurosci.* 31, 1155–1172. doi: 10.1162/jocn_a_01334

Noel, J.-P., Stevenson, R. A., and Wallace, M. T. (2018b). Atypical audiovisual temporal function in autism and schizophrenia: Similar phenotype, different cause. *Eur. J. Neurosci.* 47, 1230–1241. doi: 10.1111/ejn.13911

O'Sullivan, A. E., Crosse, M. J., Liberto, G. M. D., de Cheveigné, A., and Lalor, E. C. (2021). Neurophysiological indices of audiovisual speech processing reveal a hierarchy of multisensory integration effects. *J. Neurosci.* 41, 4991–5003. doi: 10.1523/JNEUROSCI.0906-20.2021

O'Sullivan, A. E., Lim, C. Y., and Lalor, E. C. (2019). Look at me when I'm talking to you: Selective attention at a multisensory cocktail party can be decoded using stimulus reconstruction and alpha power modulations. *Eur. J. Neurosci.* 50, 3282–3295. doi: 10.1111/ejn.14425

Odegaard, B., Wozny, D. R., and Shams, L. (2016). The effects of selective and divided attention on sensory precision and integration. *Neurosci. Lett.* 614, 24–28. doi: 10.1016/j.neulet.2015.12.039

Okamoto, H., Stracke, H., Wolters, C. H., Schmael, F., and Pantev, C. (2007). Attention improves population-level frequency tuning in human auditory cortex. *J. Neurosci.* 27, 10383–10390. doi: 10.1523/JNEUROSCI.2963-07.2007

Olasagasti, I., Bouton, S., and Giraud, A.-L. (2015). Prediction across sensory modalities: A neurocomputational model of the McGurk effect. *Cortex* 68, 61–75. doi: 10.1016/j.cortex.2015.04.008

Paré, M., Richler, R. C., ten Hove, M., and Munhall, K. G. (2003). Gaze behavior in audiovisual speech perception: The influence of ocular fixations on the McGurk effect. *Percept. Psychophys.* 65, 553–567. doi: 10.3758/bf03194582

Parker, J. L., and Robinson, C. W. (2018). Changes in multisensory integration across the life span. *Psychol. Aging* 33, 545–558. doi: 10.1037/pag0000244

Pessoa, L., Kastner, S., and Ungerleider, L. G. (2003). Neuroimaging studies of attention: From modulation of sensory processing to top-down control. *J. Neurosci.* 23, 3990–3998. doi: 10.1523/JNEUROSCI.23-10-03990.2003

Peter, M. G., Porada, D. K., Regenbogen, C., Olsson, M. J., and Lundström, J. N. (2019). Sensory loss enhances multisensory integration performance. *Cortex* 120, 116–130. doi: 10.1016/j.cortex.2019.06.003

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., and Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends Cogn. Sci.* 12, 106–113. doi: 10.1016/j.tics.2008.01.002

Schwartz, J.-L., Tiippana, K., and Andersen, T. (2010). *Disentangling Unisensory from Fusion Effects in the Attentional Modulation of McGurk Effects: A Bayesian Modeling Study Suggests That Fusion is Attention-Dependent*. Lyon: HAL.

Senkowski, D., Talsma, D., Herrmann, C. S., and Woldorff, M. G. (2005). Multisensory processing and oscillatory gamma responses: Effects of spatial selective attention. *Exp. Brain Res.* 166, 411–426. doi: 10.1007/s00221-005-2381-z

Shams, L., Ma, W. J., and Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *Neuroreport* 16, 1923–1927. doi: 10.1097/01.wnr.0000187634.68504.bb

Soto-Faraco, S., and Alsius, A. (2009). Deconstructing the McGurk-MacDonald illusion. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 580–587. doi: 10.1037/a0013483

Stein, B. E., Stanford, T. R., Ramachandran, R., Perrault, T. J., and Rowland, B. A. (2009). Challenges in quantifying multisensory integration: Alternative criteria, models, and inverse effectiveness. *Exp. Brain Res.* 198, 113–126. doi: 10.1007/s00221-009-1880-8

Stevenson, R. A., Baum, S. H., Segers, M., Ferber, S., Barense, M. D., and Wallace, M. T. (2017). Multisensory speech perception in autism spectrum disorder: From phoneme to whole-word perception. *Autism Res.* 10, 1280–1290. doi: 10.1002/aur.1776

Stolte, M., Bahrami, B., and Lavie, N. (2014). High perceptual load leads to both reduced gain and broader orientation tuning. *J. Vis.* 14:9. doi: 10.1167/14.3.9

Talsma, D., Doty, T. J., and Woldorff, M. G. (2007). Selective attention and audiovisual integration: Is attending to both modalities a prerequisite for early integration? *Cereb. Cortex* 17, 679–690. doi: 10.1093/cercor/bhk016

Talsma, D., Senkowski, D., Soto-Faraco, S., and Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends Cogn. Sci.* 14, 400–410. doi: 10.1016/j.tics.2010.06.008

Talsma, D., and Woldorff, M. G. (2005). Selective attention and multisensory integration: Multiple phases of effects on the evoked brain activity. *J. Cogn. Neurosci.* 17, 1098–1114. doi: 10.1162/0898929054475172

Tang, X., Wu, J., and Shen, Y. (2016). The interactions of multisensory integration with endogenous and exogenous attention. *Neurosci. Biobehav. Rev.* 61, 208–224. doi: 10.1016/j.neubiorev.2015.11.002

van Laarhoven, T., Stekelenburg, J. J., and Vroomen, J. (2019). Increased sub-clinical levels of autistic traits are associated with reduced multisensory integration of audiovisual speech. *Sci. Rep.* 9:9535. doi: 10.1038/s41598-019-46084-0

Wahn, B., Schmitz, L., Kingstone, A., and Böckler-Raettig, A. (2021). When eyes beat lips: Speaker gaze affects audiovisual integration in the McGurk illusion. *Psychol. Res.* 86, 1930–1943. doi: 10.1007/s00426-021-01618-y

Ward, N., Menta, A., Ulichney, V., Raileanu, C., Wooten, T., Hussey, E. K., et al. (2021). The specificity of cognitive-motor dual-task interference on balance in young and older adults. *Front. Aging Neurosci.* 13:804936. doi: 10.3389/fnagi.2021.804936

Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., et al. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proc. Natl. Acad. Sci. U.S.A.* 90, 8722–8726. doi: 10.1073/pnas.90.18.8722

Zhang, S., Xu, W., Zhu, Y., Tian, E., and Kong, W. (2020). Impaired multisensory integration predisposes the elderly people to fall: A systematic review. *Front. Neurosci.* 14:411. doi: 10.3389/fnins.2020.00411

Zion Golumbic, E., Cogan, G. B., Schroeder, C. E., and Poeppel, D. (2013). Visual input enhances selective speech envelope tracking in auditory cortex at a "cocktail party". *J. Neurosci.* 33, 1417–1426. doi: 10.1523/JNEUROSCI.3675-12.2013

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., et al. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron* 77, 980–991. doi: 10.1016/j.neuron.2012.12.037

frontiers | Frontiers in Human Neuroscience

TYPE Original Research
PUBLISHED 26 January 2023
DOI 10.3389/fnhum.2022.1058093

OPEN ACCESS

EDITED BY
Elena Nava,
University of Milano-Bicocca, Italy

REVIEWED BY
Giulia Cappagli,
Italian Institute of Technology (IIT), Italy
Dominik Osinski,
Norwegian University of Science
and Technology, Norway

*CORRESPONDENCE
Amber Maimon
✉ amber.maimon@runi.ac.il

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

SPECIALTY SECTION
This article was submitted to Sensory Neuroscience, a section of the journal Frontiers in Human Neuroscience

RECEIVED 30 September 2022
ACCEPTED 13 December 2022
PUBLISHED 26 January 2023

CITATION
Maimon A, Wald IY, Ben Oz M, Codron S, Netzer O, Heimler B and Amedi A (2023) The Topo-Speech sensory substitution system as a method of conveying spatial information to the blind and vision impaired.
Front. Hum. Neurosci. 16:1058093.
doi: 10.3389/fnhum.2022.1058093

The Topo-Speech sensory substitution system as a method of conveying spatial information to the blind and vision impaired

Amber Maimon[1,2*†], Iddo Yehoshua Wald[1,2†], Meshi Ben Oz[1,2], Sophie Codron[1,2], Ophir Netzer[3], Benedetta Heimler[4] and Amir Amedi[1,2]

[1]Baruch Ivcher School of Psychology, The Baruch Ivcher Institute for Brain, Cognition, and Technology, Reichman University, Herzliya, Israel, [2]The Ruth and Meir Rosenthal Brain Imaging Center, Reichman University, Herzliya, Israel, [3]Gonda Brain Research Center, Bar Ilan University, Ramat Gan, Israel, [4]Center of Advanced Technologies in Rehabilitation (CATR), Sheba Medical Center, Ramat Gan, Israel

Humans, like most animals, integrate sensory input in the brain from different sensory modalities. Yet humans are distinct in their ability to grasp symbolic input, which is interpreted into a cognitive mental representation of the world. This representation merges with external sensory input, providing modality integration of a different sort. This study evaluates the Topo-Speech algorithm in the blind and visually impaired. The system provides spatial information about the external world by applying sensory substitution alongside symbolic representations in a manner that corresponds with the unique way our brains acquire and process information. This is done by conveying spatial information, customarily acquired through vision, through the auditory channel, in a combination of sensory (auditory) features and symbolic language (named/spoken) features. The Topo-Speech sweeps the visual scene or image and represents objects' identity by employing naming in a spoken word and simultaneously conveying the objects' location by mapping the x-axis of the visual scene or image to the time it is announced and the y-axis by mapping the location to the pitch of the voice. This proof of concept study primarily explores the practical applicability of this approach in 22 visually impaired and blind individuals. The findings showed that individuals from both populations could effectively interpret and use the algorithm after a single training session. The blind showed an accuracy of 74.45%, while the visually impaired had an average accuracy of 72.74%. These results are comparable to those of the sighted, as shown in previous research, with all participants above chance level. As such, we demonstrate practically how aspects of spatial information can be transmitted through non-visual channels. To complement the findings, we weigh in on debates concerning models of spatial knowledge (the persistent, cumulative, or convergent models) and the capacity for spatial representation in the blind. We suggest the present study's findings support the convergence model and the scenario that posits the blind are capable of some aspects of spatial representation as depicted by the algorithm comparable to those of the sighted. Finally, we present possible future developments, implementations, and use cases for the system as an aid for the blind and visually impaired.

KEYWORDS

sensory substitution, spatial perception, sensory substitution device (SSD), blind and visually impaired people, sensory development, sensory perception

1. Introduction

Vision is commonly accepted to be the principal mediator between the objective world around us and the representation of what we perceptually experience (Cattaneo and Vecchi, 2011; Hutmacher, 2019). Visual input is known to be so dominant that it heavily influences the manner in which our other senses are processed (Posner et al., 1976), as is exhibited by well-known illusions such as the McGurk (McGurk and MacDonald, 1976) and ventriloquist effects (Bruns, 2019). Concerning spatial perception in particular, vision is considered to be especially important in forming spatial representations (Gori et al., 2014, 2020). Forming spatial representations involves acquiring a holistic image of objects and information concerning their distance, locations, and orientations relative to one's self (Struiksma et al., 2009), information classically thought to be reliably acquired predominantly through vision and insufficiently conveyed through the other senses such as audition and touch (Battal et al., 2020). Despite this, vision and audition specifically are known to be the main routes for perceiving extra-personal space, with other senses, such as the tactile sense, being associated mainly with peri-personal space and the area surrounding one's body (Van der Stoep et al., 2017).

The visually impaired and the blind gather information about their environments through multiple channels of the remaining senses. Philosopher Diderot's letter on the blind for the use of those who can see depicts this as follows: "The man-born-blind of Puiseaux works out how close he is to the fire by how hot it is, how full a receptacle is by the sound liquid makes as he decants it, and how near he is to other bodies by the way the air feels on his face (Tunstall, 2011, p. 177)." Despite the general dominance of vision, the visually impaired and the blind are known to compensate for their lack of the sense of vision by utilization of the other senses (Röder et al., 2004; Bauer et al., 2017).

As far back as biblical times, the blind used canes, similar to what we now call "white canes," as an aid in localizing and spatial orientation within their surroundings (Strong, 2009). However, not all of the blind population use a cane regularly (Blindness Statistics, n.d.). Furthermore, canes are nearly not employed at all to aid the visually impaired (Blindness Statistics, n.d.), who rely on a combination of their existing/residual vision.

A method employed by the blind specifically for acquiring spatial information is echolocation. Echolocation specifically allows for acquiring spatial representations in silent conditions, in contrast to relying on auditory cues for acquiring information from the surroundings. Echolocation, colloquially attributed to bats and dolphins in the wild, is also used by some of the blind population in a similar manner. Human echolocators make clicking sounds with their tongues and carefully listen to the echoes reverberating back to them from the objects in their surroundings. New technologies incorporate an element of color into echolocation-inspired devices, such as the EyeMusic (Abboud et al., 2014) and the Colorphone (Bizoń-Angov et al., 2021), which also incorporates a dimension of depth.

Neuroscientific findings indicate that blind expert echolocators activate the visual cortex when echolocating, specifically MT +, an area considered to be correlated with the perception of visual motion in the sighted (Thaler and Goodale, 2016). In addition, it has been shown that the sounds of echoes bouncing off of different objects activate the lateral occipital cortex, an area specifically related to shape processing, mainly through the visual perception of objects, but research conducted by our lab has shown that this area is multisensory in that it is also activated for shape processing when the information is conveyed through the tactile modality (Amedi et al., 2001, 2007). In both blind and sighted trained echolocators, a major factor underlying the ability to perform localization of objects using echolocation successfully is the element of pitch (Schenkman and Nilsson, 2011). As such, it can be understood that pitch is important for conveying spatial information. In the Topo-Speech algorithm, differences in pitch represent different locations on the y-axis, which may make the algorithm more intuitive, though this warrants future exploration as it is possible that pitch in echolocation might be helpful due to the reflection characteristics of the objects, while no such phenomena is explored here.

Braille reading and spoken language have also both been correlated with visual cortex activation (Sadato et al., 1996, 1998; Büchel et al., 1998; Seydell-Greenwald et al., 2021). The Braille reading method, invented in 1824, can be considered one of the earliest sensory substitution methods (Ptito et al., 2021). Braille conveys verbal information through haptic or tactile stimulation (Kristjánsson et al., 2016). Braille readers must use extreme accuracy and sensitivity to discriminate between patterns of raised dots with their fingers and translate this code into meaningful semantic information (Hamilton and Pascual-Leone, 1998). This indicates that language can serve as a potential substitute in the absence of vision.

It has been suggested (Röder et al., 2004) that when vision is unavailable, the following three scenarios are possible: The first scenario posits that there will be a lack or decrease in the sensory capabilities due to the lack of an essential sense (vision) (Gori et al., 2014; Cappagli et al., 2017b; Martolini et al., 2020a). The second is that no difference will be observed (Haber et al., 1993; Morrongiello et al., 1995). While the third suggests that there will be compensation, defined as better performance or surpassing the capabilities seen in the sighted, due to other sensory mechanisms making up for the lack of the visual sense (Roder et al., 1999; Amedi et al., 2003, 2004 Collignon et al., 2006; Sabourin et al., 2022).

This is currently a central debate, with different bodies of evidence supporting the various hypotheses (Röder et al., 2004). In this study, we ask which of these scenarios is better supported. Moreover, we explore a specific kind of spatial perception induced by the Topo-Speech system in the visually impaired, a group underrepresented in the wealth of literature that mainly compares the sighted and the blind. We explored the difference (or lack thereof) in the performance of the visually impaired vs. the blind when using this system for conveying a certain kind of spatial perception. In addition, we explore the performance of the visually impaired population with the system as well. Exploring the abilities of the visually impaired in this case could serve as a particularly interesting intermediate in that their sensory development with respect to vision is distinct from that of both the sighted and the blind.

Recently we developed a novel sensory substitution algorithm in our lab that combines verbal and spatial information (Heimler et al., 2019; Netzer et al., 2021). The Topo-Speech algorithm used in the current study represents the verbal naming of an object in a way that conveys its location in space such that in the vertical axis, objects located higher are represented by a higher pitch and lower by a lower pitch. The horizontal axis is mapped temporally from left to right, such that the closer the object is to the left, the sooner one hears the stimulus. This representation provides the user with information that allows them to simultaneously know both the identity of the objects and their locations in space by correlating the spoken word (for identity) with defined auditory characteristics (for location).

Prior research has shown that sighted individuals can successfully learn to use this algorithm for identifying spatial positions after undergoing a single training session and even generalize to locations they had not been trained on (Netzer et al., 2021). Yet thus far, research exploring this system has provided a first proof of concept focused on the technical aspects of the Topo-Speech algorithm and the ability of sighted blindfolded participants. This study on the sighted showed that they were able to understand and use this system with success levels well above chance level. This study expands upon these findings to explore the applicability of this approach, and its advantages and disadvantages, in visually impaired and blind individuals.

The key, primary, goal of this present study is practical—to extend the prior research to assess the ability of the visually impaired and blind to understand the algorithm and explore whether their modified visual experience throughout life influenced their ability to perform with the system. The study aims to provide a proof of concept for the possible future development of the algorithm as an aid for the visually impaired. In addition, we demonstrate how some aspects of spatial information can be transmitted through non-visual channels. More specifically, the correlation between a sensory method for conveying spatial information through audition, and a symbolic one, for conveying object identity through language. We suggest that this is of particular significance because it allows for taking the high-complexity visual data and translating it to symbolic representation, alongside lower bandwidth data to a sensory representation.

Another complementary goal of this study is more theoretical. While there is very little dispute concerning the dominance of vision in sensory perception and in forming our holistic representation of the world (Cattaneo and Vecchi, 2011; Cappagli and Gori, 2019; Hutmacher, 2019), we know that the human brain (as the root of how we perceive the world around us) is exceedingly capable of adaptation to its circumstances and forthcomings.

These matters are pertinent due to the differential neurodevelopment in the visually impaired/blind as compared to the sighted and the distinct experiences of the different populations with respect to forming spatial representations throughout their lives. This is relevant for another debate of whether the deficit (insofar as there is one) concerning spatial perception is indeed perceptual or cognitive (Bleau et al., 2022). For example, it is known that the human brain is structured such that it can compensate by way of other senses. There are two main strategies for this, one it the "taking over" of visual function by an increase in the efficiency of other functions (for a review, see Bedny, 2017), and the other is through sensory substitution (for a review, see Maidenbaum et al., 2014a). Sensory substitution is the transfer of information commonly provided through one sense through an alternate sense. The sensory substitution method used in the current study uses a sweep line technique, whose use in sensory substitution owes its beginning to the vOICe (Amedi and Meijer, 2005) sensory substitution device (SSD) that introduced an algorithm that scans the visual scene from left to right. It translates it into sounds using spectrographic sound synthesis and other audio enhancement techniques, pixel by pixel. The corresponding series of sounds is known as a "soundscape," in which the horizontal axis is represented by the time of presentation and panning, while the vertical axis corresponds to tone frequency, and the level of intensity (loudness) of the sound represents brightness (Striem-Amit et al., 2012). Other SSDs employing the sweep line technique incorporate additional dimensions of the visual scene in the transformation, such as the EyeMusic, which represents a color dimension through different timbres of sound in a musical pentatonic scale (Abboud et al., 2014). We will inquire into the question of which mechanism is taking place by employing the method of sensory substitution of spatial information commonly acquired by vision in the sighted through auditory properties.

Yet another powerful mediator between the world and our perception of it is language. Throughout the history of mankind, spoken language has served as a distinguishing feature of humans from other species. Language is considered so powerful that it was thought to threaten god's supremacy in the story of the tower of Babel. As such, it is no surprise that our brains are very much attuned to language processing. In the brain, language plays such a significant role that the visual deprivation in the blind sparks neuroplastic mechanisms which enable higher cognitive functions, such as language processing, to activate the visual cortex (Amedi et al., 2004; Merabet et al., 2005; Bedny, 2017) alongside specific spatial language processing (Struiksma et al., 2011).

Furthermore, research indicates that language provides a central means for acquiring spatial information in the blind (Afonso et al., 2010). Though symbolic and not sensory, language has been shown to bring about spatial representations to the same extent as perceptual auditory information (Loomis et al., 2002). Following these insights concerning the significance of language, our lab has previously developed the Topo-Speech algorithm that conveys, *via* spoken language, object identity *via* spoken language (Heimler et al., 2019; Netzer et al., 2021). As such, the current study is not only practical, but we lay the groundwork for further research exploring the perception of space *via* language and sensory information of those with no visual experience.

2. Materials and methods

2.1. Participants

Twenty two adults (10 Female) participated in the study, eight of whom are blind and 14 visually impaired (**Table 1**). Blind participants in the study were determined *via* a certificate of blindness. A person is entitled to receive the certificate if they are totally legally blind or with visual acuity of 3/60 m (Ministry of welfare and social security, n.d.). The visually impaired do not have a blind certificate, so their status as visually impaired was verified according to their entitlement to obtaining a driving license. As defined by national regulation, to be eligible for a driver's license, one must present with visual acuity between 6/12 and 6/6 m (Assuta-optic, n.d.). Those who do not have visual acuity in this range, even after vision correction methods such as glasses, are defined as visually impaired. Visually impaired participants were blindfolded during the experiment. When speaking of the sighted, we are comparing to the participants in Netzer et al. (2021). Their participants were 15 sighted adults (nine women; aged 27.2 ± 1.57 years). All participants had no known hearing/balance impairments or neurological conditions. All participants were above the age of 18 with a mean age of (35.64 years ± 10.75 years). None of the participants had prior experience with either the Topo-Speech algorithm nor any other sensory substitution device. This study received full ethics approval from the Reichman University Institutional Review Board (IRB). All participants received monetary compensation of 80 shekels per hour for their participation in the experiment alongside reimbursement for their transportation to and from the university.

TABLE 1 Participant information.

#	Age	Visual impairment	Details
1	35	Vision impaired	Retinopathy of prematurity, nystagmus
2	41	Vision impaired	Retinitis pigmentosa
3	59	Late blind	Retinitis pigmentosa, glaucoma, blind since age 44
4	22	Vision impaired	Glaucoma
5	57	Vision impaired	Retinitis pigmentosa
6	29	Early blind	Glaucoma, Blind since age 15
7	29	Congenitally blind	
8	38	Congenitally blind	
9	31	Congenitally blind	
10	37	Vision impaired	Retinitis pigmentosa
11	38	Congenitally blind	
12	26	Vision impaired	Photophobia
13	43	Vision impaired	Stickler syndrome
14	44	Vision impaired	Optic atrophy
15	28	Vision impaired	Albinism
16	26	Vision impaired	Septo-optic dysplasia/de-morsier
17	24	Vision impaired	Retinal degeneration
18	52	Vision impaired	Cataract, nystagmus
19	25	Vision impaired	Albinism, astigmatism, nystagmus
20	42	Vision impaired	Cone dystrophy
21	23	Congenitally blind	
22	35	Congenitally blind	

2.2. Experiment design and algorithm

The Topo-Speech algorithm is a sweep line algorithm that scans the visual scene from left to right. The x axis of the

TABLE 2 Words used in the experiment.

English translation	English transliteration	Word in Hebrew
Words used in the training stage of the experiment		
Earring	Agil	עגיל
Window	Chalon	חלון
Wand	Sharvit	שרביט
Flower	Perah	פרח
Hat	Kova	כובע
Chair	Kise	כיסא
Newspaper	Iton	עיתון
Doll	Buba	בובה
Flag	Degel	דגל
Carrot	Gezer	גזר
Box	Argaz	ארגז
Branch	Anaf	ענף
Pitcher	Kankan	קנקן
Shoe	Naal	נעל
Balloon	Balon	בלון
Pot	Atzitz	עציץ
Soap	Sabon	סבון
Sock	Gerev	גרב
Boots	Magaf	מגף
Words used in the testing stage of the experiment		
Mug	Sefel	ספל
Ruler	Sargel	סרגל
Lighter	Matzit	מצית
Wallet	Arnak	ארנק
Bottle	Bakbuk	בקבוק
Ball	Kadur	כדור
Hammer	Patish	פטיש
Knife	Sakin	סכין
Notepad	Pincas	פנקס
Shirt	Chultza	חולצה
Computer	Machshev	מחשב
Flute	Chalil	חליל
Lemon	Limon	לימון
Fork	Mazleg	מזלג
Teaspoon	Capit	כפית
Orange	Tapuz	תפוז
Kettle	Kumkum	קומקום
Glove	K'fafa	כפפה
Leaf	Aleh	עלה
Binder	Klasser	קלסר
Box	Kufsa	קופסה
Lock	Manul	מנעול
Cable	Chevel	חבל
Bobby Pin	Sika	סיכה
Stapler	Shadchan	שדכן
Pan	Machvat	מחבת
Letter	Michtav	מכתב
Paintbrush	Mik'hol	מכחול
Button	Kaftor	כפתור

visual scene or image is mapped to time and represents the objects' horizontal locations, while the y-axis of the visual scene or image is mapped to the pitch of the soundscape. As such, the algorithm functions in a manner such that if one hears word 1 followed temporally by word 2, then corresponding object 1 was located further to the left of the visual scene than corresponding object 2 (representing the x axis). If one hears word 1 higher in pitch than word 2, then corresponding object 1 is located higher in the visual scene than corresponding object 2. The content of the words represents the identity of the object scanned (for example shoe, book). For the purpose of this study, a database of 60 highly frequent words in the Hebrew language was professionally recorded, after which the words were modified for the format of the Topo-Speech algorithm and trials using the Audacity audio editing software. The training stage consisted of 27 trials in total. During training, trials were presented in a random order, where each of the nine possible locations was tested three times, and a word could not appear twice in the same location. The testing stage consisted of 90 trials in total, with words appearing in each location 10 times. The words used were all matched to two syllables for consistency and represented objects with no inherent spatial content, such as "na-al" (shoe) and "se-fer" (book) instead of "sky" and "carpet" which may be associated with upper and lower parts of space, respectively. See Table 2 for the complete list of words. Two short consecutive beeps signified each word presentation's start and end points.

A 3 × 3 grid was created for the experiment, with the dimensions 80 × 80 cm. The grid was hung on the wall at 160 cm. A digital beep indicated the beginning of the word presentation, followed by a word stating the object's identity presented after various delays from the initial beep. A second beep indicated the end of the word presentation, thus defining the borders of the x-axis. The spatial location of each word could be presented in one of 3 different pitches in the y-axis: Pitches Low C, Low A#, and Middle G#, and each trial lasted a total of 2 s.

Before beginning the training, participants received a general explanation about the concept of sensory substitution devices: "Sensory substitution devices (SSDs) are algorithms that convey visual information *via* other sensory modalities, in this case, audition. The algorithm does this substitution based on two principles: First, the algorithm uses language to identify the objects, and second, it uses their location in space."

The participants were instructed to reach their hands forward and freely feel the grid. Meanwhile, they were told that the algorithm maps a space of three rows and three columns, creating a 3 × 3 grid. Then, participants were explained the rules by which they would locate the objects during the experiment as follows: "The X-axis is mapped to the time domain (from left to right). The participants hear a beep at the beginning and end of each word (and there is a difference between them so that you know when it is the beginning and when it is the end). The

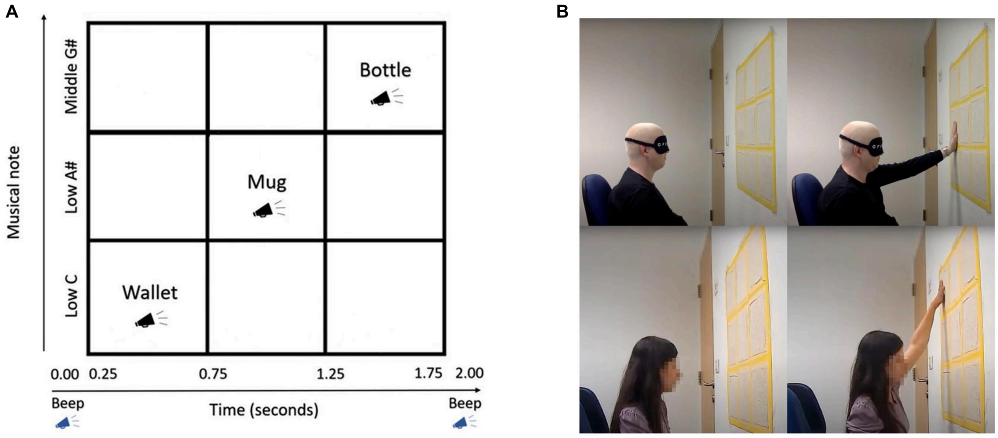


FIGURE 1
(A) Illustration of the Topo-Speech algorithm as employed in the current study (adapted from Netzer et al., 2021) **(B)** Experimental set-up for visually impaired (top) and blind (bottom).

closer a stimulus is heard to one of the beeps, the participant will be able to perceive its proximity to the left or right ends of the grid. The y-axis is mapped to the pitch, so the higher the participant hears a word, the higher the corresponding stimulus will be placed on the grid."

Participants underwent two stages during the study, a training stage and then a testing stage. They were asked to listen carefully to the auditory stimuli, after which they were instructed to touch the cell on the 3 × 3 grid that they thought represented the location of the object presented through audition. The participants touching of the grid represents sensorimotor compatibility to show that the auditory information can become spatial, and connects to the field with the help of the sensorimotor action of reaching out hands. Sending the hand to the location indicates that the person can perform this conversion. From an applicative point of view, the purpose of the algorithm is to help the blind and visually impaired to operate in the world, and when we are in the environment we don't call out the names of objects to get them. Rather, we want the algorithm to tell us where the object is in space so that we can easily reach out and take it. Thus, the feedback is motor and not verbal.

During training the participants would listen to each word presentation, and receive feedback on their response with regard to the spatial location represented. They would receive feedback on whether their response was correct or incorrect, and if their answer was incorrect they were directed to the location of the correct answer in each training trial. They could repeat the playback of the word presentation as many times as desired, and if the response was incorrect, the same word was repeated. The training stage lasted for an average of 15:36 ± 9.06 min [mean ± standard deviation (SD)], with 14 min for visually impaired and 17 min for the blind. During the testing stage, consisting of 90 trials, each word was presented twice, after which a choice was made by the participant. No feedback was provided during this stage. The participant responses were recorded. Following every 30 trials, the participant was offered a 2-min break, which they could choose to not take if they wished. With regards to removing the blindfold during the break, visually impaired participants were turned around so as not to see the grid. The testing stage lasted for an average of 23:27 ± 5:34 min (mean ± *SD*).

Figure 1 illustrates the experimental set-up, the 3 × 3 grid, and depicts how the participants chose their answer during the experiment. The top two images are of a blindfolded visually impaired participant, the bottom two are images of a blind participant.

2.3. Statistical analyses

Because the sample sizes in both groups were relatively small (less than 50 participants), non-parametric statistical tests were used, including the Mann–Whitney to compare success rates between groups and the Wilcoxon signed-rank test (equivalent to the *t*-test) for independent samples comparisons, specifically the participants' performance to chance level. These tests were performed using SPSS Statistics 25. Cumulative success rates and the average number of mistakes per 10 trials were calculated and shown below in **Figure 2**. The cumulative success rate is a variable calculated for each subject. For each subject, instead of calculating all of the successes on all the stages of the test at once, each time, a cumulative success rate is considered up to the current trial in order to assess the learning curve of the subjects. As such, we calculate the cumulative percentage of success the participant had up to a certain trial and divide it by the trial number.

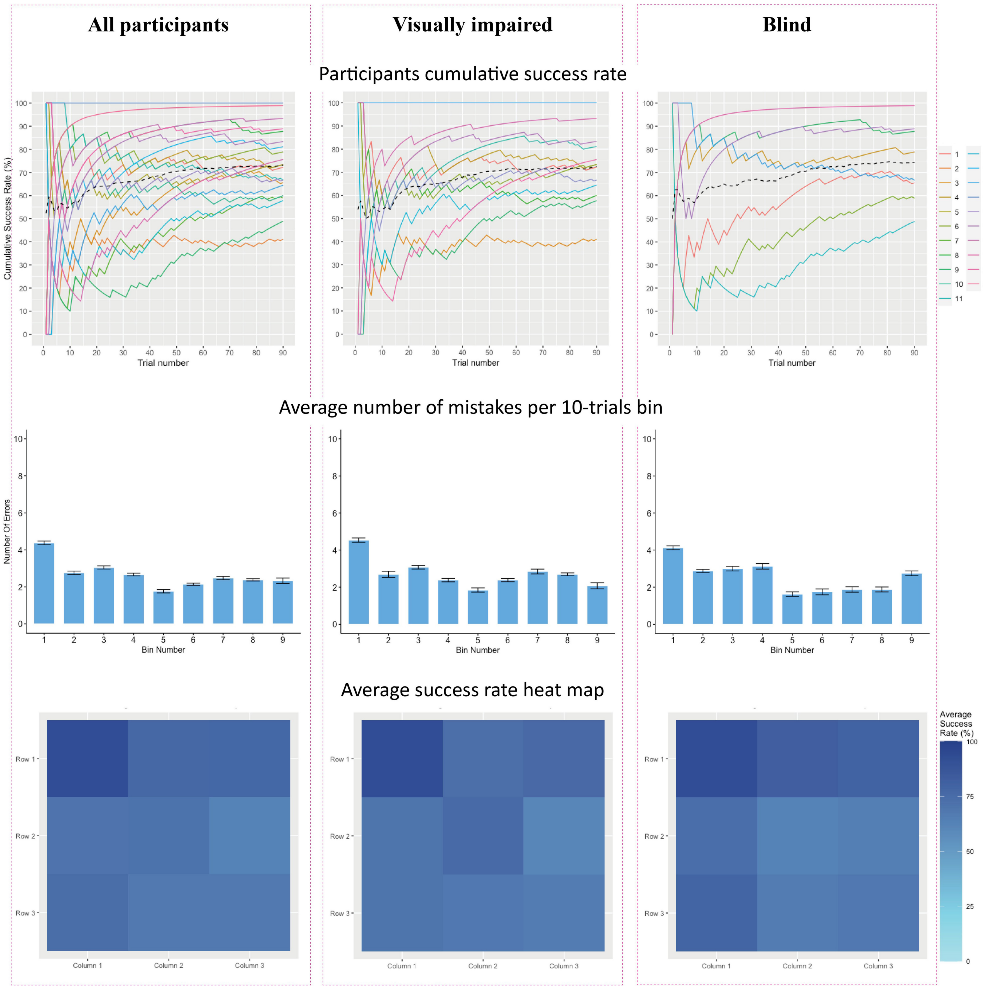


FIGURE 2

(Top row)- Cumulative success rates by trial number. Each participant is represented by a colored line, with the group average represented by a black dashed line. **(Middle row)** The average number of mistakes per 10 trials. **(Bottom row)** of figures portrays a heat map of the average success rate. Darker shades represent higher rates of success. **(Left column)** results across all participants (n = 21). **(Middle column)** Average success rates for blind participants (n = 8). **(Right column)** Average success rates for visually impaired participants (n = 13).

3. Results

3.1. Both groups successfully learned to use the algorithm following a short training period

The group average and SD during the experiment were: 73.39% ± 15.89%. One visually impaired participant was removed due to his being extremely uncooperative during the experimental procedure. Table 3 specifies each participant's success rates and standard deviation during the testing stage. Table 4 summarizes the success rates and SD of the individual participants divided into blind and visually impaired. The performance of all participants was greater than chance.

Figure 3 specifies the success rate of each individual participant compared to the chance level at 11%. As can be observed in the figure, all participants performed above chance level. This was confirmed with a one-sample Wilcoxon signed-rank test, indicating that the participants performed significantly better than chance level; $T = 231.00$, $z = 4.015$, $p < 0.001$.

3.2. The performance of the blind was not significantly different from that of the visually impaired

A Mann-Whitney test indicated no significant differences between group success rates, $U(N_{\text{blind}} = 8, N_{\text{visually impaired}} = 13) = 49.500$, $z = -0.181$, $p = 0.860$.

TABLE 3 Individual success rate and SD for each participant.

Subject number	Success rate (%)	SD (%)
1	72.22	45.04
2	41.11	49.48
3	65.56	47.78
4	73.33	44.47
5	73.33	44.47
6	78.89	41.04
7	58.89	49.48
8	87.78	32.94
9	48.89	50.27
10	60.00	49.26
11	66.67	47.40
12	57.78	49.67
13	82.22	38.45
14	64.44	48.14
15	100.00	0.00
16	66.67	47.40
17	83.33	37.48
18	95.56	20.72
19	75.56	43.22
20	90.00	30.17
21	98.89	10.54

TABLE 4 Summary of success rates of each group.

Impairment	Number of subjects	Success rate (%): mean (*SD*)
Blind	8	74.45 (± 17.20)
Visually impaired	13	72.74 (± 15.72)

3.3. The performance of the visually impaired and the blind was not significantly different from that of the sighted

A Mann-Whitney test revealed no significant difference between success rates in the present sample compared to a previous study that used sighted participants (Netzer et al., 2021); $U(N_{\text{Non-sighted}} = 21, N_{\text{sighted}} = 14) = 101.00$, $z = -1.550$, $p = 0.127$.

In order to analyze whether participants' performance improved throughout the experiment, cumulative success rates for each participant were modeled. The top row of **Figure 2** represents the result for each participant across all 90 trials, along with the group's average cumulative success rate. All three graphs show a slight positive gradient, indicating improvement in success rate across the experiment, with a plateau from trial 50 onward. Another method to model participants' learning during the experiment is by assessing the number of mistakes made by participants during the experiment. This is depicted in the middle row. The experiment was divided into nine bins of ten trials each, in which the average number of participant errors was calculated for every 10 trials.

To represent the success rates across participants, heat maps were created, where the average success rate for each square on the 3 × 3 grid is represented with a cell in the graph. The first graph on the bottom row highlights that while all participants had very high success rates, they identified the top-left cell with the highest accuracy and were least accurate in identifying the middle-right cell. The bottom middle and bottom right graphs present heat maps of blind and visually impaired participants separately. Visually impaired participants identified the top left the most accurately, along with top-right and middle closely after, whereas blind participants revealed a clear advantage in identifying the top row, as well as on the left, with top-left and bottom-left having the highest success rate.

4. Discussion

Prior research conducted in our lab showed that sighted blindfolded participants could learn to use the Topo-Speech algorithm with an accuracy of 80.24 percent (Netzer et al., 2021). In this study, we expanded upon this by evaluating for the first time this system's ability to convey spatial information to the population it was aimed at: the blind and visually impaired. The current study results show that both the blind and visually impaired were capable of learning to use the algorithm with success rates comparable to those of the sighted, with no statistically significant difference in their performance. The blind showed an accuracy of 74.45 percent, while the visually impaired had an average accuracy of 72.74. It is apparent that the blind performed better than the visually impaired, though not significantly better. The same trend held for training, with the blind having a shorter training time than the visually impaired but a longer training time than that shown in the (blindfolded) sighted participants in our previous study, though the difference was not found to be statistically significant (averaging ∼10 min for sighted, 14 min for visually impaired and 17 min for the blind). As such, the main goal of this study—to test whether the system is highly intuitive also to fully blind or visually impaired—has been positively achieved. In this study, we were primarily interested in assessing the system's feasibility and practical functionality for the blind and visually impaired. As such, a limitation to be noted concerning the comparisons to the sighted is that the groups were not matched for age and gender. Future studies could perform a direct comparison between these groups as well.

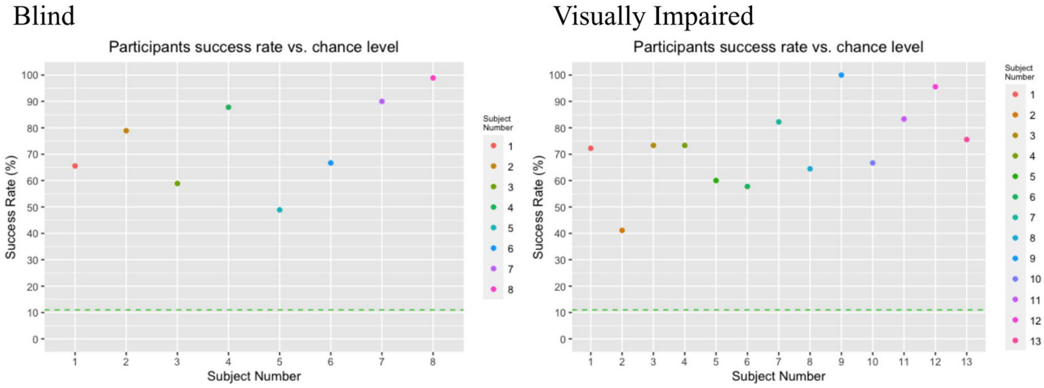


FIGURE 3
Individual participant success rate vs. chance level. Each subject is represented by a filled dot, and the chance level is represented by a dotted line, at position 11.00 on the y axis. The chance level is 11% success rate due to a correct answer being one of nine possible locations.

4.1. Spatial perception and representation when vision is lacking

Concerning the more theoretical debate—a central debate with relation to blindness, congenital blindness, in particular, is the ability or inability of the blind to reach the same or a similar level of skill as the sighted with respect to numerous tasks. This question is particularly interesting with relation to tasks in which vision is classically thought to play a central role, for example, spatial perception. On the one hand, it was commonly accepted that the blind and visually impaired have a significant impairment with respect to their sense of space and capability for forming spatial representations (Gori et al., 2014; Cappagli et al., 2017b; Martolini et al., 2020a). However, research now indicates that following dedicated training, the blind can become more capable of spatial localization (Gaunet et al., 1997; Cappagli et al., 2017a). It is now thought that the blind can perform spatial tasks with the same level of ability as the sighted when the information is delivered as auditory or tactile input, and some research even indicates that they may reach better performance (Roder et al., 1999; Collignon et al., 2006). In addition, three lifelong models represent trajectories of obtaining a spatial understanding in the blind. Two models suggest that vision is such a crucial element in spatial knowledge that the blind devoid of vision, are at an insurmountable disadvantage. The persistent model states that the blind have an initial disadvantage compared to the sighted, which persists throughout life. The cumulative model posits that the disability not only persists but even leads to an increase in the discrepancy between the abilities of the blind and the sighted (the blind improve in their abilities over time while the blind do not). The convergent model, on the other hand, suggests that the blind have an initial disadvantage with respect to spatial knowledge, and yet this "converges" with the abilities of the sighted as a result of experience throughout life and training, whether explicit or implicit (Schinazi et al., 2016; Aggius-Vella et al., 2017; Finocchietti et al., 2017; Cuppone et al., 2018, 2019; Cappagli et al., 2019; Martolini et al., 2020b, 2022). It is now known that the blind brain provides for compensatory mechanisms for lack of vision and visual deprivation from numerous studies showing activation of the visual cortex in response to various spatial tasks (Striem-Amit et al., 2012; Abboud et al., 2015 and also reviews by Kupers and Ptito, 2014; Maidenbaum et al., 2014a; Ricciardi et al., 2014). The findings of this study support the convergence model alongside the scenario that posits that the ability of the blind and the visually impaired to understand a subset of spatial representation delivered by the auditory information provided by the system is not inferior. Moreover, the wealth of research showing improved performance in the blind population after training corresponds to this direction of the convergence model, therefore agreeing with and supporting our hypotheses.

The current study also strengthens research indicating that while the blind are capable of some aspects of spatial perception to a similar extent as the sighted, the visually impaired show (a non-significant in our case) trend for slightly poorer performance. A previous study showed that individuals with low residual vision (peripheral) were less capable of sound localization and performed worse than both their blind and sighted counterparts (Lessard et al., 1998). Another study showed that children with visual impairments are less capable of updating spatial coordinates as compared to the sighted (Martolini et al., 2020a). It could be that in the case of the visually impaired, compensatory neuroplasticity takes place to a lesser extent than in the blind, and yet their vision is severely impaired in comparison to the sighted. On the other hand, Cappagli et al. (2017b) found that while the blind perform more poorly on tasks related to spatial hearing, the visually impaired perform at the same level as their normally sighted peers.

In this study, alongside the objective qualitative experiment, the participants were also asked several subjective questions related to spatial perception and their user experience with the training and algorithm during the experiment. Their responses, though anecdotal, emphasize some interesting points. When asked, "How do you get information about the locations of objects in your environment in everyday life?" The visually impaired showed an automatic tendency to lean on residual vision, even when it is limited in capacity, and other senses are fully functional and intact. This strengthens the hegemony of the visual modality with respect to spatial perception, as all but one of the visually impaired indicated vision as their main mode of acquiring a spatial representation of objects in their surroundings. For example, one participant replied, "My vision is good enough for me to manage without relying on my hearing." Another stated, "I see, and when I can't, then I use hearing." These subjective reports support the interpretation expressed by Cattaneo and Vecchi (2011), who ask, "why is vision so important in our life?" The answer is quite pragmatic: because the visual is "easy." On the other hand, all but two of the blind participants, who cannot default to vision, reported not only using hearing, but many specifically reported that they rely on "asking" other people, thereby acquiring the information through language. One participant said: "If it's an environment I know, I know everything where it is, and if I'm in a new environment, I don't know it, then I have to ask and study it first and then I move on" and another responded: "By asking people or arranging the objects in an order that I can choose." This comes alongside hearing, as expressed by another participant: "According to the sound mainly and later I make a map in the head of the structure and everything." These reports further support the understanding that due to their reliance on their (limited) visual capabilities as opposed to other wholly intact sensory modalities, compensatory neuroplasticity is less likely to take place, and if it does, to a lesser degree than in the blind. Even more so, it is likely that their defaulting to vision has a detrimental effect on their ability to compensate behaviorally by way of other senses, as is exhibited by the Colavita effect (Colavita, 1974; Spence et al., 2011).

4.2. Conveying spatial information through non-visual channels

While our prior research provided a proof of concept for the general usability of the algorithm as a method of conveying spatial information through language, the current findings take this one step forward. We believe this study serves as a proof of concept for using language-based sensory substitution systems such as the Topo-Speech to aid the visually impaired and the blind.

Some tools developed for the blind and visually impaired tackle the issue of spatial localization from the practical perspective, designed, for example, to allow the blind to gather information from their surroundings specifically pertaining to distances, navigation, and obstacle detection for a particular aim, such as independent mobility. One such tool, the EyeCane, is an electronic travel aid (ETA) that relies on multiple sensory stimuli. The EyeCane, for example, integrates auditory cues with haptic ones allowing the user to identify objects and barriers in their surroundings by manipulations in the frequency of the multisensory cues (Chebat et al., 2011, 2015; Buchs et al., 2017; Maidenbaum et al., 2014b). For an extensive review of other technologies developed for assisting the blind, see Ptito et al. (2021). A particularly interesting avenue for further research with this algorithm related to the integration of haptic and auditory cues could be the association between Braille reading (or lack thereof) and success with using the Topo-Speech algorithm. All of the blind participants in this study were Braille readers.

Considering these findings, we speculate that Braille readers may have higher success rates and find the Topo-Speech algorithm more intuitive than non-Braille readers. This is further supported by the correspondence between Braille reading, which is from left to right, and the Topo-Speech in which a word presented temporally closer to the beginning of the stimulus presentation is closer to the left side of the "visual field."

Braille reading also conveys language information through haptic, or tactile, stimulation (Kristjánsson et al., 2016), as Braille readers must use extreme accuracy and sensitivity to discriminate between patterns of raised dots with their fingers and translate this code into meaningful semantic information. Various studies have shown that Braille readers have an enlarged sensory representation of the reading finger, compared to sighted and blind non-Braille-readers, through recording somatosensory evoked potentials (Pascual-Leone and Torres, 1993). Mapping the motor cortical areas that represent reading fingers through transcranial magnetic stimulation has revealed that this enlargement is also seen in blind Braille readers (Pascual-Leone et al., 1993). A plausible explanation of these research findings is that the afferent information extracted by blind Braille readers from their fingerpad may be more detailed and specific, causing them to succeed in the discriminatory task that is Braille reading (Hamilton and Pascual-Leone, 1998).

Braille reading also connects to another interesting finding of this study. Compared to the visually impaired and the sighted, the blind participants showed a clearer tendency to correct answers on the left side of the answer grid. Research indicates that spatial orientation and processing is lateralized to the right hemisphere in the blind (Rinaldi et al., 2020) as well as in the sighted (Vogel et al., 2003). This right hemisphere lateralization has been shown to be more prominent and substantial in the blind (Rinaldi et al., 2020). One of the mechanisms correlated with this is Braille, written from left to right (Rinaldi et al., 2020) (similar to the stimuli presented in the present experiment: a word closer to the first beep occurs on the left side of the visual field or closer to zero on the X-axis). Nevertheless, this warrants further research, as an alternative explanation for this could be more trivially related to the ability to perceive time exactly. This could be tested by exploring a change in the direction of the sweep.

Braille represents a language system composed of symbols, the Braille letters. The current study demonstrates the potential of using mixed methods for conveying information through sensory substitution coupled with symbolic means. Such mixed representations of information would use sensory substitution, in which the information is conveyed perceptually, alongside symbolic information, representing more complex combinations of information. In this case, location properties *via* audio features and spoken language that provide symbols our brains are attuned to processing. Training with purely visual to auditory sensory substitution devices (such as the EyeMusic upon which the Topo-Speech algorithm is based) can take tens of hours of training. On the other hand, the Topo-Speech algorithm can successfully be trained in short training sessions of under 20 min in all populations tested. We suggest that this is partly due to the combination as it allows for taking the high-complexity visual data and translating it to a symbolic representation and the scalable, lower bandwidth data to a sensory representation. This also serves to strengthen the interpretation of the brain as a task-selective, sensory independent organ. Under this interpretation, different brain areas are correlated with tasks (such as perceiving the 3D shape of objects) rather than senses (such as vision).

4.3. Future directions and implementations

To further refine and establish the findings of this work, we find it valuable to explore several adaptations of the algorithm and evaluation method. Particularly, we would wish to experiment with changing the direction of the sweep, the mapping of pitch ranges to height, and different time intervals. Such experiments could eliminate any biases caused due to the specific choice of one of these factors, as well as help in understanding their role in changing perception. Hence guiding future work in optimizing such representation algorithms.

With regard to expanding on the experimental design, a direction that we have previously explored in the sighted and could provide another meaningful evaluation of the blind and visually impaired would be to extend the spatial representation provided by the algorithm to the backward space. While vision in the sighted has a limited range of 210 horizontal degrees (Strasburger and Pöppel, 2002), audition spans the full 360 degrees. As indicated by our prior research with the Eyemusic algorithm (Shvadron et al., under review) and the Topo-Speech (Heimler et al., 2019; Netzer et al., 2021), it is possible to convey information from the back spatial field by way of audition. Our future implementation of the Topo-Speech algorithm could explore the feasibility and implications of such expansion of the spatial field in the blind and visually impaired, who are not susceptible to a limited range of vision in the front space to begin with.

Another possible adaptation of the Topo-Speech algorithm could incorporate a tactile element as well. We have previously shown how coupling tactile information to auditory speech using a "speech to touch" SSD enhances auditory speech comprehension (Cieśla et al., 2019, 2022). Adding this sensory modality to the algorithm may enhance its effectiveness by means of coupling tactile feedback to the auditory stimuli or adding more information, such as the representation of another dimension, such as depth *via* vibration intensity or frequency. This is further supported by an abundant body of research indicating that the blind show similar to better performance than the sighted, particularly in auditory and tactile tasks (Lessard et al., 1998; Van Boven et al., 2000; Gougoux et al., 2004, 2005; Voss et al., 2004; Doucet et al., 2005; Collignon et al., 2008).

Going forward, we aim to expand the capabilities of the existing algorithm beyond its current limitations by accounting for the lack of ability to represent objects simultaneously, offering a more continuous representation of the space, as well as more dimensional information describing a scene. For example, a future implementation of the Topo-Speech algorithm could convey dimensions such as depth through different manipulations such as volume and more or through other sensory stimuli, as suggested above. Another such "dimension" could be one of color—it is now known that the blind have a concept of color though historically thought to be "ungraspable" to those who have never experienced it (Kim et al., 2021) and that colors can affect spatial perception for dimensions and size (Oberfeld and Hecht, 2011; Yildirim et al., 2011). This dimension of perception, not currently available to the blind, could be conveyed similarly to the one used in the EyeMusic algorithm, developed in our lab by using different timbres of sound.

This study also provides a stepping stone toward fMRI studies of the sighted, the blind, and the visually impaired using the Topo-Speech algorithm. Aside from the activation of language areas such as Broca's area and auditory areas, we would expect to see activation in the visual cortex of the blind. It would be of particular interest to compare the visual cortex activation in the blind to that in the visually impaired and the sighted when performing the task when blindfolded. Such a study could possibly shed light on the differences between the blind and the visually impaired concerning compensation by way of neuroplasticity and the extent thereof. In addition, we would be interested in seeing whether there are areas activated specifically for the combination between audition and language with respect to spatial perception.

When implemented in different systems, the Topo-speech algorithm could have several practical use cases representing two general categories, set and non-set scene implementations. Set scenes are ones where the information to be portrayed is fixed and not dynamically changing. An example of such scenarios would be using the Topo-Speech algorithm in a system that provides information to the blind/visually impaired concerning emergency exits in buildings, information about items in a museum, or elements in virtual reality situations. Systems designed for use in non-set scenes (in which objects and their locations change dynamically), could incorporate real-time artificial intelligence, for example, using image recognition to identify the objects to be named by the algorithm. The incorporation of artificial intelligence could open a wealth of possibilities for the blind and visually impaired with respect to providing them with freedom and independence in unfamiliar or changing environments. Artificial intelligence is already being integrated into rehabilitative systems for the sensory impaired, for example retinal prostheses (Barnes, 2012; Weiland et al., 2012) and hearing aids (Crowson et al., 2020; Lesica et al., 2021), yet SSD systems hold the potential for providing a more transparent and automatic perceptual experience (Ward and Meijer, 2010; Maimon et al., 2022) and therefore could be particularly powerful when combined with real time computer vision.

In addition, as the content represented by the algorithm can theoretically be adapted at will, one can imagine different operational modalities, that could even be alternated between by the user to match their needs, or automatically according to different use cases. For example, a previous study conducted by our institute has shown the feasibility of adding a "zooming in" element for increasing resolution when using a visual to auditory sensory substitution device in the blind (Buchs et al., 2016). An advanced implementation of the Topo-Speech could incorporate a "zooming in" feature to increase the resolution from "fruit" to "banana," "apple," or a "zooming out" feature to allow for general contextualization for example "home" or "gym." These features can be particularly useful to the blind for independence and navigation in space. In addition, the algorithm could be attuned by the user to specific contextual categories, such as navigation elements (elevator, stairs) or people located spatially within a scene (supermarket crowded with people and objects), identification of the age or gender of people in different spatial locations of the scene, and more.

Data availability statement

The original contributions presented in this study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by the Reichman University Institutional Review Board (IRB). The patients/participants provided their written informed consent to participate in this study.

Author contributions

AM: writing—original draft, writing—review and editing, methodology, conceptualization, visualization, project administration, and supervision. IW: writing—original draft, writing—review and editing, methodology, conceptualization, and visualization. MB: writing—original draft, investigation, and formal analysis. SC: writing—original draft, investigation, visualization, and formal analysis. ON and BH: methodology, conceptualization, and investigation. AA: writing—original draft and writing—review and editing, project administration, supervision, resources, conceptualization, investigation, methodology, funding acquisition. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the ERC Consolidator Grant (773121 NovelExperiSense) and the Horizon GuestXR (101017884) grant—both to AA.

Acknowledgments

We would like to thank Shaqed Tzabbar for her assistance with running participants and analysis of questionnaire responses.

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

References

Abboud, S., Hanassy, S., Levy-Tzedek, S., Maidenbaum, S., and Amedi, A. (2014). EyeMusic: Introducing a "visual" colorful experience for the blind using auditory sensory substitution. *Restor. Neurol. Neurosci.* 32, 247–257. doi: 10.3233/RNN-130338

Abboud, S., Maidenbaum, S., Dehaene, S., and Amedi, A. (2015). A number-form area in the blind. *Nat. Commun.* 6:6026. doi: 10.1038/ncomms7026

Afonso, A., Blum, A., Katz, B. F., Tarroux, P., Borst, G., and Denis, M. (2010). Structural properties of spatial representations in blind people: Scanning images constructed from haptic exploration or from locomotion in a 3-D audio virtual environment. *Memory Cogn.* 38, 591–604. doi: 10.3758/MC.38.5.591

Aggius-Vella, E., Campus, C., Finocchietti, S., and Gori, M. (2017). Audio motor training at the foot level improves space representation. *Front. Integr. Neurosci.* 11:36. doi: 10.3389/fnint.2017.00036

Amedi, A., and Meijer, P. (2005). Neural correlates of visual-to-auditory sensory substitution in proficient blind users.

Amedi, A., Floel, A., Knecht, S., Zohary, E., and Cohen, L. G. (2004). Transcranial magnetic stimulation of the occipital pole interferes with verbal processing in blind subjects. *Nat. Neurosci.* 7, 1266–1270. doi: 10.1038/nn1328

Amedi, A., Malach, R., Hendler, T., Peled, S., and Zohary, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. *Nat. Neurosci.* 4, 324–330. doi: 10.1038/85201

Amedi, A., Raz, N., Pianka, P., Malach, R., and Zohary, E. (2003). Early 'visual'cortex activation correlates with superior verbal memory performance in the blind. *Nat. Neurosci.* 6, 758–766. doi: 10.1038/nn1072

Amedi, A., Stern, W. M., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., et al. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nat. Neurosci.* 10, 687–689. doi: 10.1038/nn1912

Assuta-optic (n.d.). *Eyes on the road: everything you need to know about an eye test for a green form*. Available online at: https://www.assuta-optic.co.il (accessed September 29, 2022).

Barnes, N. (2012). The role of computer vision in prosthetic vision. *Image Vis. Comput.* 30, 478–479. doi: 10.1016/j.imavis.2012.05.007

Battal, C., Occelli, V., Bertonati, G., Falagiarda, F., and Collignon, O. (2020). General enhancement of spatial hearing in congenitally blind people. *Psychol. Sci.* 31, 1129–1139. doi: 10.1177/0956797620935584

Bauer, C. M., Hirsch, G. V., Zajac, L., Koo, B. B., Collignon, O., and Merabet, L. B. (2017). Multimodal MR-imaging reveals large-scale structural and functional connectivity changes in profound early blindness. *PLoS One* 12:e0173064. doi: 10.1371/journal.pone.0173064

Bedny, M. (2017). Evidence from blindness for a cognitively pluripotent cortex. *Trends Cogn. Sci.* 21, 637–648. doi: 10.1016/j.tics.2017.06.003

Bizoń-Angov, P., Osiński, D., Wierzchoń, M., and Konieczny, J. (2021). Visual echolocation concept for the colorophone sensory substitution device using virtual reality. *Sensors* 21:237. doi: 10.3390/s21010237

Bleau, M., Paré, S., Chebat, D. R., Kupers, R., Nemargut, J. P., and Ptito, M. (2022). Neural substrates of spatial processing and navigation in blindness: An ALE meta-analysis. *Front. Neurosci.* 16:1010354. doi: 10.3389/fnins.2022.1010354

Blindness Statistics (n.d.). *National Federation of the Blind*. Available online at: https://nfb.org/resources/blindness-statistics (accessed September 28, 2022).

Bruns, P. (2019). The ventriloquist illusion as a tool to study multisensory processing: An update. *Front. Integr. Neurosci.* 13:51. doi: 10.3389/fnint.2019.00051

Büchel, C., Price, C., Frackowiak, R. S., and Friston, K. (1998). Different activation patterns in the visual cortex of late and congenitally blind subjects. *Brain* 121, 409–419. doi: 10.1093/brain/121.3.409

Buchs, G., Maidenbaum, S., Levy-Tzedek, S., and Amedi, A. (2016). Integration and binding in rehabilitative sensory substitution: Increasing resolution using a new Zooming-in approach. *Restor. Neurol. Neurosci.* 34, 97–105. doi: 10.3233/RNN-150592

Buchs, G., Simon, N., Maidenbaum, S., and Amedi, A. (2017). Waist-up protection for blind individuals using the EyeCane as a primary and secondary mobility aid. *Restor. Neurol. Neurosci.* 35, 225–235. doi: 10.3233/RNN-160686

Cappagli, G., and Gori, M. (2019). "The role of vision on spatial competence," in *Visual impairment and blindness-what we know and what we have to know*, (London: IntechOpen). doi: 10.5772/intechopen.89273

Cappagli, G., Finocchietti, S., Baud-Bovy, G., Cocchi, E., and Gori, M. (2017a). Multisensory rehabilitation training improves spatial perception in totally but not partially visually deprived children. *Front. Integr. Neurosci.* 11:29. doi: 10.3389/fnint.2017.00029

Cappagli, G., Finocchietti, S., Cocchi, E., and Gori, M. (2017b). The impact of early visual deprivation on spatial hearing: A comparison between totally and partially visually deprived children. *Front. Psychol.* 8:467. doi: 10.3389/fpsyg.2017.00467

Cappagli, G., Finocchietti, S., Cocchi, E., Giammari, G., Zumiani, R., Cuppone, A. V., et al. (2019). Audio motor training improves mobility and spatial cognition in visually impaired children. *Sci. Rep.* 9:3303. doi: 10.1038/s41598-019-39981-x

Cattaneo, Z., and Vecchi, T. (2011). *Blind vision: the neuroscience of visual impairment*. Cambridge, MA: MIT press. doi: 10.7551/mitpress/9780262015035.001.0001

Chebat, D. R., Maidenbaum, S., and Amedi, A. (2015). Navigation using sensory substitution in real and virtual mazes. *PLoS One* 10:e0126307. doi: 10.1371/journal.pone.0126307

Chebat, D. R., Schneider, F. C., Kupers, R., and Ptito, M. (2011). Navigation with a sensory substitution device in congenitally blind individuals. *Neuroreport* 22, 342–347. doi: 10.1097/WNR.0b013e3283462def

Cieśla, K., Wolak, T., Lorens, A., Heimler, B., Skarżyński, H., and Amedi, A. (2019). Immediate improvement of speech-in-noise perception through multisensory stimulation via an auditory to tactile sensory substitution. *Restor. Neurol. Neurosci.* 37, 155–166. doi: 10.3233/RNN-190898

Cieśla, K., Wolak, T., Lorens, A., Mentzel, M., Skarżyński, H., and Amedi, A. (2022). Effects of training and using an audio-tactile sensory substitution device on speech-in-noise understanding. *Sci. Rep.* 12:3206. doi: 10.1038/s41598-022-06855-8

Colavita, F. B. (1974). Human sensory dominance. *Percept. Psychophys.* 16, 409–412. doi: 10.3758/BF03203962

Collignon, O., Renier, L., Bruyer, R., Tranduy, D., and Veraart, C. (2006). Improved selective and divided spatial attention in early blind subjects. *Brain Res.* 1075, 175–182. doi: 10.1016/j.brainres.2005.12.079

Collignon, O., Voss, P., Lassonde, M., and Lepore, F. (2008). Cross-modal plasticity for the spatial processing of sounds in visually deprived subjects. *Exp. Brain Res.* 192, 343–358. doi: 10.1007/s00221-008-1553-z

Crowson, M. G., Lin, V., Chen, J. M., and Chan, T. C. (2020). Machine learning and cochlear implantation—a structured review of opportunities and challenges. *Otol. Neurotol.* 41, e36–e45. doi: 10.1097/MAO.0000000000002440

Cuppone, A. V., Cappagli, G., and Gori, M. (2018). Audio feedback associated with body movement enhances audio and somatosensory spatial representation. *Front. Integr. Neurosci.* 12:37. doi: 10.3389/fnint.2018.00037

Cuppone, A. V., Cappagli, G., and Gori, M. (2019). Audio-motor training enhances auditory and proprioceptive functions in the blind adult. *Front. Neurosci.* 13:1272. doi: 10.3389/fnins.2019.01272

Doucet, M. E., Guillemot, J. P., Lassonde, M., Gagné, J. P., Leclerc, C., and Lepore, F. (2005). Blind subjects process auditory spectral cues more efficiently than sighted individuals. *Exp. Brain Res.* 160, 194–202. doi: 10.1007/s00221-004-2000-4

Finocchietti, S., Cappagli, G., and Gori, M. (2017). Auditory spatial recalibration in congenital blind individuals. *Front. Neurosci.* 11:76. doi: 10.3389/fnins.2017.00076

Gaunet, F., Martinez, J. L., and Thinus-Blanc, C. (1997). Early-blind subjects' spatial representation of manipulatory space: Exploratory strategies and reaction to change. *Perception* 26, 345–366. doi: 10.1068/p260345

Gori, M., Amadeo, M. B., and Campus, C. (2020). Spatial metric in blindness: Behavioural and cortical processing. *Neurosci. Biobehav. Rev.* 109, 54–62. doi: 10.1016/j.neubiorev.2019.12.031

Gori, M., Sandini, G., Martinoli, C., and Burr, D. C. (2014). Impairment of auditory spatial localization in congenitally blind human subjects. *Brain* 137, 288–293. doi: 10.1093/brain/awt311

Gougoux, F., Lepore, F., Lassonde, M., Voss, P., Zatorre, R. J., and Belin, P. (2004). Pitch discrimination in the early blind. *Nature* 430, 309–309. doi: 10.1038/430309a

Gougoux, F., Zatorre, R. J., Lassonde, M., Voss, P., and Lepore, F. (2005). A functional neuroimaging study of sound localization: Visual cortex activity predicts performance in early-blind individuals. *PLoS Biol.* 3:e27. doi: 10.1371/journal.pbio.0030027

Haber, R. N., Haber, L. R., Levin, C. A., and Hollyfield, R. (1993). Properties of spatial representations: Data from sighted and blind subjects. *Percept. Psychophys.* 54, 1–13. doi: 10.3758/BF03206932

Hamilton, R. H., and Pascual-Leone, A. (1998). Cortical plasticity associated with Braille learning. *Trends Cogn. Sci.* 2, 168–174. doi: 10.1016/S1364-6613(98)01172-3

Heimler, B., Shur, A., Netzer, O., Behor, T., and Amedi, A. (2019). "The topo-speech algorithm: an intuitive sensory substitution for spatial information," in *Proceedings of the 2019 international conference on virtual rehabilitation (ICVR)*, (Tel Aviv: IEEE), 1–2. doi: 10.1109/ICVR46560.2019.8994658

Hutmacher, F. (2019). Why is there so much more research on vision than on any other sensory modality? *Front. Psychol.* 10:2246. doi: 10.3389/fpsyg.2019.02246

Kim, J. S., Aheimer, B., Montané Manrara, V., and Bedny, M. (2021). Shared understanding of color among sighted and blind adults. *Proc. Natl Acad. Sci. U.S.A.* 118:e2020192118. doi: 10.1073/pnas.2020192118

Kristjánsson, Á, Moldoveanu, A., Jóhannesson, ÓI., Balan, O., Spagnol, S., Valgeirsdóttir, V. V., et al. (2016). Designing sensory-substitution devices: Principles, pitfalls and potential 1. *Restor. Neurol. Neurosci.* 34, 769–787. doi: 10.3233/RNN-160647

Kupers, R., and Ptito, M. (2014). Compensatory plasticity and cross-modal reorganization following early visual deprivation. *Neurosci. Biobehav. Rev.* 41, 36–52. doi: 10.1016/j.neubiorev.2013.08.001

Lesica, N. A., Mehta, N., Manjaly, J. G., Deng, L., Wilson, B. S., and Zeng, F. G. (2021). Harnessing the power of artificial intelligence to transform hearing healthcare and research. *Nat. Mach. Intellig.* 3, 840–849. doi: 10.1038/s42256-021-00394-z

Lessard, N., Paré, M., Lepore, F., and Lassonde, M. (1998). Early-blind human subjects localize sound sources better than sighted subjects. *Nature* 395, 278–280. doi: 10.1038/26228

Loomis, J. M., Lippa, Y., Klatzky, R. L., and Golledge, R. G. (2002). Spatial updating of locations specified by 3-d sound and spatial language. *J. Exp. Psychol. Learn. Mem. Cogn.* 28:335. doi: 10.1037/0278-7393.28.2.335

Maidenbaum, S., Abboud, S., and Amedi, A. (2014a). Sensory substitution: Closing the gap between basic research and widespread practical visual rehabilitation. *Neurosci. Biobehav. Rev.* 41, 3–15. doi: 10.1016/j.neubiorev.2013.11.007

Maidenbaum, S., Hanassy, S., Abboud, S., Buchs, G., Chebat, D. R., Levy-Tzedek, S., et al. (2014b). The "EyeCane", a new electronic travel aid for the blind: Technology, behavior & swift learning. *Restor. Neurol. Neurosci.* 32, 813–824. doi: 10.3233/RNN-130351

Maimon, A., Yizhar, O., Buchs, G., Heimler, B., and Amedi, A. (2022). A case study in phenomenology of visual experience with retinal prosthesis versus visual-to-auditory sensory substitution. *Neuropsychologia* 173:108305. doi: 10.1016/j.neuropsychologia.2022.108305

Martolini, C., Amadeo, M. B., Campus, C., Cappagli, G., and Gori, M. (2022). Effects of audio-motor training on spatial representations in long-term late blindness. *Neuropsychologia* 176:108391. doi: 10.1016/j.neuropsychologia.2022.108391

Martolini, C., Cappagli, G., Campus, C., and Gori, M. (2020a). Shape recognition with sounds: Improvement in sighted individuals after audio–motor training. *Multisens. Res.* 33, 417–431. doi: 10.1163/22134808-20191460

Martolini, C., Cappagli, G., Luparia, A., Signorini, S., and Gori, M. (2020b). The impact of vision loss on allocentric spatial coding. *Front. Neurosci.* 14:565. doi: 10.3389/fnins.2020.00565

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0

Merabet, L. B., Rizzo, J. F., Amedi, A., Somers, D. C., and Pascual-Leone, A. (2005). What blindness can tell us about seeing again: Merging neuroplasticity and neuroprostheses. *Nat. Rev. Neurosci.* 6, 71–77. doi: 10.1038/nrn1586

Ministry of welfare and social security, (n.d.). *About blindness and visual impairments*. Available online at: https://www.gov.il/he/departments/guides/molsa-people-with-disabilities-people-with-blindness-and-visual-impairments-about?chapterIndex=2 (accessed September 29, 2022).

Morrongiello, B. A., Timney, B., Humphrey, G. K., Anderson, S., and Skory, C. (1995). Spatial knowledge in blind and sighted children. *J. Exp. Child Psychol.* 59, 211–233. doi: 10.1006/jecp.1995.1010

Netzer, O., Heimler, B., Shur, A., Behor, T., and Amedi, A. (2021). Backward spatial perception can be augmented through a novel visual-to-auditory sensory substitution algorithm. *Sci. Rep.* 11:11944. doi: 10.1038/s41598-021-88595-9

Oberfeld, D., and Hecht, H. (2011). Fashion versus perception: The impact of surface lightness on the perceived dimensions of interior space. *Hum. Fact.* 53, 284–298. doi: 10.1177/0018720811407331

Pascual-Leone, A., and Torres, F. (1993). Plasticity of the sensorimotor cortex representation of the reading finger in Braille readers. *Brain* 116, 39–52. doi: 10.1093/brain/116.1.39

Pascual-Leone, A., Cammarota, A., Wassermann, E. M., Brasil-Neto, J. P., Cohen, L. G., and Hallett, M. (1993). Modulation of motor cortical outputs to the reading hand of braille readers. *Ann. Neurol.* 34, 33–37. doi: 10.1002/ana.410340108

Posner, M. I., Nissen, M. J., and Klein, M. (1976). Visual dominance:An information processing account of its origins and significance. *Psychol. Rev.* 83, 157–171. doi: 10.1037/0033-295X.83.2.157

Ptito, M., Bleau, M., Djerourou, I., Paré, S., Schneider, F. C., and Chebat, D. R. (2021). Brain-machine interfaces to assist the blind. *Front. Hum. Neurosci.* 15:638887. doi: 10.3389/fnhum.2021.638887

Ricciardi, E., Bonino, D., Pellegrini, S., and Pietrini, P. (2014). Mind the blind brain to understand the sighted one! Is there a supramodal cortical functional architecture? *Neurosci. Biobehav. Rev.* 41, 64–77.

Rinaldi, L., Ciricugno, A., Merabet, L. B., Vecchi, T., and Cattaneo, Z. (2020). The effect of blindness on spatial asymmetries. *Brain Sci.* 10:662. doi: 10.3390/brainsci10100662

Röder, B., Rösler, F., and Spence, C. (2004). Early vsion impairs tactile perception in the blind. *Curr. Biol.* 14, 121–124. doi: 10.1016/j.cub.2003.12.054

Roder, B., Teder-SaÈlejaÈrvi, W., Sterr, A., RoÈsler, F., Hillyard, S. A., and Neville, H. J. (1999). Improved auditory spatial tuning in blind humans. *Nature* 400, 162–166. doi: 10.1038/22106

Sabourin, C. J., Merrikhi, Y., and Lomber, S. G. (2022). Do blind people hear better? *Trends Cogn. Sci.* 26, 999-1012. doi: 10.1016/j.tics.2022.08.016

Sadato, N., Pascual-Leone, A., Grafman, J., Deiber, M. P., Ibanez, V., and Hallett, M. (1998). Neural networks for Braille reading by the blind. *Brain* 121, 1213–1229. doi: 10.1093/brain/121.7.1213

Sadato, N., Pascual-Leone, A., Grafman, J., Ibañez, V., Deiber, M. P., Dold, G., et al. (1996). Activation of the primary visual cortex by Braille reading in blind subjects. *Nature* 380, 526–528. doi: 10.1038/380526a0

Schenkman, B. N., and Nilsson, M. E. (2011). Human echolocation: Pitch versus loudness information. *Perception* 40, 840–852. doi: 10.1068/p6898

Schinazi, V. R., Thrash, T., and Chebat, D. R. (2016). Spatial navigation by congenitally blind individuals. *WIREs Cogn. Sci.* 7, 37–58. doi: 10.1002/wcs.1375

Seydell-Greenwald, A., Wang, X., Newport, E., Bi, Y., and Striem-Amit, E. (2021). Spoken language comprehension activates the primary visual cortex. *bioRxiv [Preprint]*. doi: 10.1101/2020.12.02.408765

Shvadron, S., Snir, A., Maimon, A., Yizhar, O., Harel, S., Poradosu, K., et al. (under review). *Shape detection beyond the visual field using a visual to auditory sensory augmentation device.*

Spence, C., Parise, C., and Chen, Y.-C. (2011). "The Colavita visual dominance effect," in *Frontiers in the neural bases of multisensory processes*, eds M. M. Murray and M. Wallace (Boca Raton, FL: CRC Press), 523–550. doi: 10.1201/9781439812174-34

Strasburger, H., and Pöppel, E. (2002). "Visual field," in *Encyclopedia of neuroscience*, eds G. S. Adelman and B. H. Smith (Amsterdam: Elsevier), 2127–2129.

Striem-Amit, E., Cohen, L., Dehaene, S., and Amedi, A. (2012). Reading with sounds: Sensory substitution selectively activates the visual word form area in the blind. *Neuron* 76, 640–652. doi: 10.1016/j.neuron.2012.08.026

Strong, P. (2009). *The history of the white cane*. Watertown, MA: Paths to Literacy.

Struiksma, M. E., Noordzij, M. L., and Postma, A. (2009). What is the link between language and spatial images? Behavioral and neural findings in blind and sighted individuals. *Acta Psychol.* 132, 145–156. doi: 10.1016/j.actpsy.2009.04.002

Struiksma, M. E., Noordzij, M. L., Neggers, S. F., Bosker, W. M., and Postma, A. (2011). Spatial language processing in the blind: Evidence for a supramodal representation and cortical reorganization. *PLoS One* 6:e24253. doi: 10.1371/journal.pone.0024253

Thaler, L., and Goodale, M. A. (2016). Echolocation in humans: An overview. *Wiley Interdiscip. Rev. Cogn. Sci.* 7, 382–393. doi: 10.1002/wcs.1408

Tunstall, K. E. (2011). *Blindness and enlightenment: an essay: with a new translation of Diderot's' letter on the Blind'and La Mothe Le Vayer's' Of a Man Born Blind'*. London: Bloomsbury Publishing.

Van Boven, R. W., Hamilton, R. H., Kauffman, T., Keenan, J. P., and Pascual–Leone, A. (2000). Tactile spatial resolution in blind Braille readers. *Neurology* 54, 2230–2236. doi: 10.1212/WNL.54.12.2230

Van der Stoep, N., Postma, A., and Nijboer, T. C. (2017). "Multisensory perception and the coding of space," in *Neuropsychology of space: spatial functions of the human brain*, eds A. Postma and I. J. M. van der Ham (London: Academic Press). doi: 10.1016/B978-0-12-801638-1.00004-5

Vogel, J. J., Bowers, C. A., and Vogel, D. S. (2003). Cerebral lateralization of spatial abilities: A meta-analysis. *Brain Cogn.* 52, 197–204.

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

Ward, J., and Meijer, P. (2010). Visual experiences in the blind induced by an auditory sensory substitution device. *Conscious. Cogn.* 19, 492–500. doi: 10.1016/j.concog.2009.10.006

Weiland, J. D., Parikh, N., Pradeep, V., and Medioni, G. (2012). Smart image processing system for retinal prosthesis. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2012, 300–303. doi: 10.1109/EMBC.2012.6345928

Yildirim, K., Hidayetoglu, M. L., and Capanoglu, A. (2011). Effects of interior colors on mood and preference: Comparisons of two living rooms. *Percept. Mot. Skills* 112, 509–524. doi: 10.2466/24.27.PMS.112.2.509-524

frontiers | Frontiers in Human Neuroscience

TYPE Perspective
PUBLISHED 02 February 2023
DOI 10.3389/fnhum.2023.1108354

OPEN ACCESS

EDITED BY
Elena Nava,
University of Milano-Bicocca, Italy

REVIEWED BY
James P. Trujillo,
Radboud University, Netherlands
Huriye Atilgan,
University of Oxford, United Kingdom
Jun-ichiro Watanabe,
Hitachi, Japan

*CORRESPONDENCE
Stefania Benetti
✉ stefania.benetti@unitn.it

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

SPECIALTY SECTION
This article was submitted to
Sensory Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 25 November 2022
ACCEPTED 11 January 2023
PUBLISHED 02 February 2023

CITATION
Benetti S, Ferrari A and Pavani F (2023)
Multimodal processing in face-to-face interactions: A bridging link between psycholinguistics and sensory neuroscience.
Front. Hum. Neurosci. 17:1108354.
doi: 10.3389/fnhum.2023.1108354

Multimodal processing in face-to-face interactions: A bridging link between psycholinguistics and sensory neuroscience

Stefania Benetti[1,2]*†, Ambra Ferrari[3]† and Francesco Pavani[1,2]

[1]Centre for Mind/Brain Sciences, University of Trento, Trento, Italy, [2]Interuniversity Research Centre "Cognition, Language, and Deafness", CIRCLeS, Catania, Italy, [3]Max Planck Institute for Psycholinguistics, Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, Netherlands

In face-to-face communication, humans are faced with multiple layers of discontinuous multimodal signals, such as head, face, hand gestures, speech and non-speech sounds, which need to be interpreted as coherent and unified communicative actions. This implies a fundamental computational challenge: optimally binding only signals belonging to the same communicative action while segregating signals that are not connected by the communicative content. How do we achieve such an extraordinary feat, reliably, and efficiently? To address this question, we need to further move the study of human communication beyond speech-centred perspectives and promote a multimodal approach combined with interdisciplinary cooperation. Accordingly, we seek to reconcile two explanatory frameworks recently proposed in psycholinguistics and sensory neuroscience into a neurocognitive model of multimodal face-to-face communication. First, we introduce a psycholinguistic framework that characterises face-to-face communication at three parallel processing levels: multiplex signals, multimodal gestalts and multilevel predictions. Second, we consider the recent proposal of a lateral neural visual pathway specifically dedicated to the dynamic aspects of social perception and reconceive it from a multimodal perspective ("lateral processing pathway"). Third, we reconcile the two frameworks into a neurocognitive model that proposes how multiplex signals, multimodal gestalts, and multilevel predictions may be implemented along the lateral processing pathway. Finally, we advocate a multimodal and multidisciplinary research approach, combining state-of-the-art imaging techniques, computational modelling and artificial intelligence for future empirical testing of our model.

KEYWORDS

multimodal communication, face-to-face interactions, social actions, lateral cortical processing pathway, psycholinguistics, sensory neuroscience

Introduction

In face-to-face communication, we encounter multiple layers of discontinuous multimodal signals: head, face, mouth movements, hand gestures, speech and non-speech sounds. This implies a fundamental computational challenge: optimally binding only signals belonging to the same communicative action while segregating unrelated signals (Noppeney, 2021). Within this challenge, the temporal misalignment of fast-changing signals across different sensory channels raises a central binding problem (Chen and Vroomen, 2013). Finally, each conversational partner is taxed by fast turn-taking dynamics (Levinson, 2016). Despite these critical constraints, we process multimodal communicative signals faster than speech alone (Holler et al., 2018; Drijvers and Holler, 2022). Crucially, we use non-verbal communicative signals to facilitate semantic understanding (Özyürek, 2014) and pragmatic inference (Holler, 2022). How do we achieve such an extraordinary feat?

To address this question, we need to move beyond the prominent speech-centred research perspective on the neurocognitive mechanisms of human communication. Building on previous calls for the need to study language in its multimodal manifestation and ecological context (Levinson and Holler, 2014; Vigliocco et al., 2014; Hasson et al., 2018; Perniss, 2018), the view we put forward here seeks to reconcile two explanatory frameworks recently proposed in psycholinguistics and sensory neuroscience. Specifically, we first highlight that verbal and non-verbal communicative signals are integrated to represent socially relevant acts (Levinson and Holler, 2014) through domain-general mechanisms of multimodal integration and prediction (Holler and Levinson, 2019). Accordingly, we then reconceive the neuroscientific evidence of a third visual pathway, specialised for dynamic aspects of social perception (Pitcher and Ungerleider, 2021), from a multimodal perspective. Finally, we propose that the resulting brain network implements the sensory processing gateway necessary toward successful multimodal processing and interpretation of face-to-face communicative signals.

Multimodal processing in face-to-face interactions: A possible computational framework

Holler and Levinson (2019) recently outlined the key computational principles that support fast and efficient multimodal processing in face-to-face communication, with the ultimate goal of interpreting communicative social actions (**Figure 1A**). First, domain-general mechanisms of multimodal integration (Stein, 2012; Noppeney, 2021) are hypothesised to be co-opted for detecting communicative signals. For example, faster processing of multimodal relative to unimodal communicative inputs mirrors multimodal facilitation outside the domain of communication in humans (Murray et al., 2001; Senkowski, 2005; Diederich et al., 2009) and animals (Gingras et al., 2009). Holler and Levinson (2019) proposed that multimodal interactions resting on statistical regularities among sensory inputs allow chunking the stream of concurrent dynamic inputs into *multiplex signals* at a perceptual, pre-semantic level. Further, the statistical regularities between multiplex signals and communicative meanings generate *multimodal gestalts* that bear semantic and pragmatic value, thus signalling a specific social action. For example, eyebrow frowns often accompany a raising voice pitch to signal the intention to ask a question (Nota et al., 2021). Mechanisms of Gestalt perception (Wagemans et al., 2012), social affordance (Gallagher, 2020), and relevance (Sperber and Wilson, 1995) may jointly contribute to the recognition of multimodal communicative gestalts (Trujillo and Holler, 2023). Finally, the recognition of a specific social action may trigger top-down *multilevel predictions* about how the message will unfold in time. For example, frowning and pointing at an object typically anticipates a question about that object, triggering top-down hierarchical predictions at multiple sensory levels (e.g., vocal sounds, bodily movements) and linguistic levels (e.g., words, sentential units). Multiplex signals, multimodal gestalts, and multilevel predictions are thought to interact in a continuous, dialectic process, leading to incremental unification while the message unfolds (Hagoort, 2005, 2019). Specifically, this supports a parallel processing framework whereby the beginning of the message simultaneously activates multiple potential interpretations (i.e., multimodal gestalts). As the message unfolds, concurrent bottom-up sensory processing and multilevel predictions iteratively refine each other toward a final gestalt solution (Trujillo and Holler, 2023). Such a parallel account accommodates evidence that processing of communicative social actions starts early (Redcay and Carlson, 2015), perhaps in parallel to semantic comprehension (Tomasello et al., 2022).

Supporting this framework, there is substantial psycholinguistic evidence for systematic associations between facial-bodily signals and social actions (Holler and Levinson, 2019; Nota et al., 2021). Moreover, the early emergence of these perceptual associations in infants (Cameron-Faulkner et al., 2015), as well as parallels in non-human primates (Rossano and Liebal, 2014), suggest they might be deeply rooted in the human onto- and phylogenesis.

Multimodal processing in face-to-face interactions: A possible neural framework

Accumulating evidence (Pitcher et al., 2014; Walbrin and Koldewyn, 2019; Landsiedel et al., 2022) suggests that dynamic visual aspects of social perception (e.g., face, hand and body movements across the visual field) cannot be easily accommodated within the classic dual-stream model for visual perception (Ungerleider and Mishkin, 1982). Accordingly, resting on both anatomical and functional evidence in humans and non-human primates, Pitcher and Ungerleider (2021) proposed the existence of a third visual processing pathway (**Figure 1B**) that projects on the lateral cortical surface from the early visual cortex into the mid-posterior superior temporal sulcus (pSTS) *via* motion-selective occipito-temporal areas (V5/hMT). Consistent evidence shows that pSTS preferentially responds to multiple types of dynamic social bodily inputs including eye, mouth, hands, and body movements (Allison et al., 2000; Hein and Knight, 2008; Deen et al., 2020). Importantly, both anterior hMT (Desimone and Ungerleider, 1986; Huk et al., 2002) and pSTS (Bruce et al., 1981; Pitcher et al., 2020; Finzi et al., 2021) respond to dynamic signals across both visual hemifields in human and non-human primates, in opposition to the contralateral field bias that characterises the ventral pathway (Finzi et al., 2021). Together, these functional properties are thought to support social interaction, which is an inherently dynamic process requiring the integration of sensory

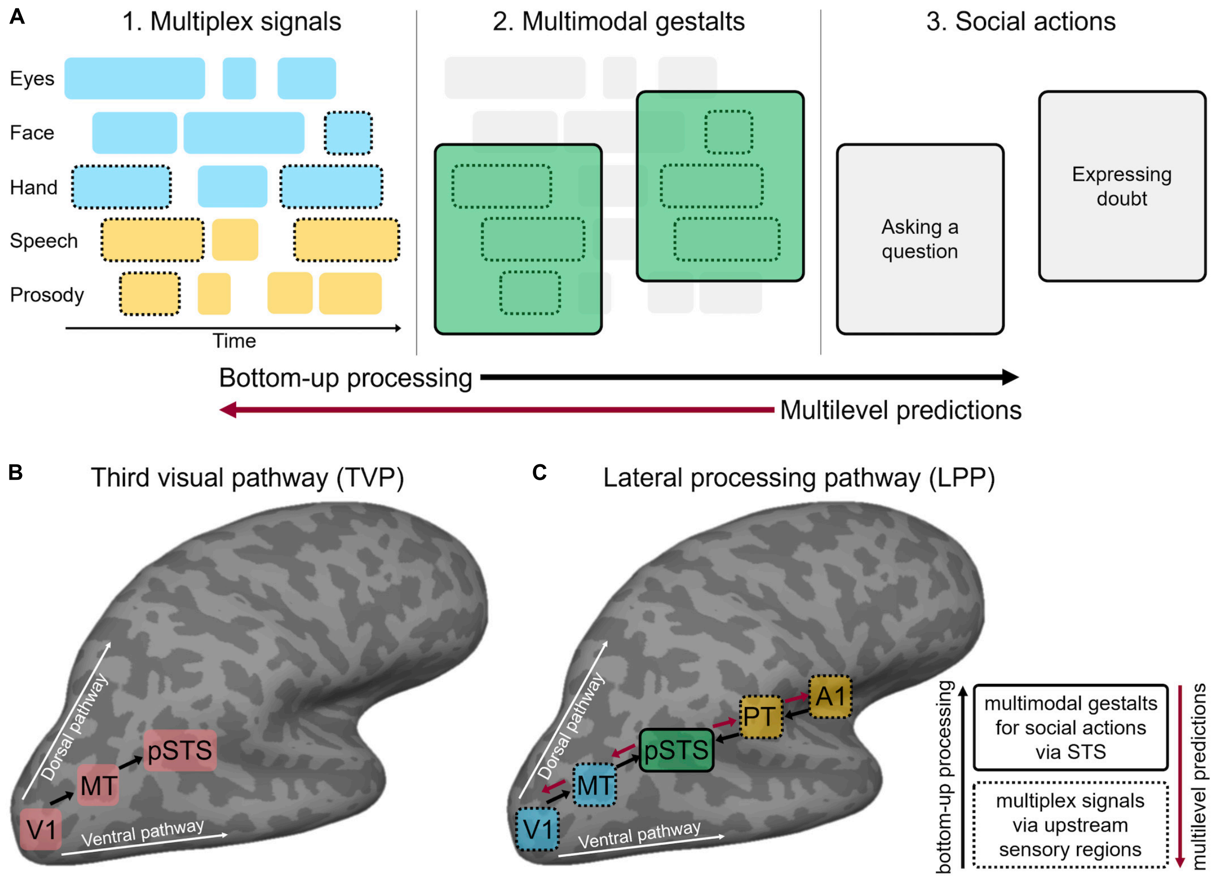


FIGURE 1
(A) Face-to-face communication comprises multiple layers of discontinuous multimodal signals emitted by different articulators (eyes, face, etc.) over time (visual in cyan, auditory in yellow). To enable effective communication, interlocutors must bind only signals belonging to the same communicative action while segregating tangential, unrelated signals that do not share the communicative content. Temporal statistical regularities allow chunking coherent communicative inputs into multiplex signals (dashed contours) at a perceptual, pre-semantic level (A1). Further, statistical regularities between multiplex signals and communicative meanings generate multimodal gestalts (solid contours) that bear semantic and pragmatic value (A2) and thus signal a specific communicative intention in conversation, i.e., social action (A3). Finally, social action recognition may trigger top-down multilevel predictions across hierarchically organised linguistic and perceptual levels. Following a parallel processing framework, concurrent bottom-up sensory processing (black arrow) and multilevel predictions (red arrow) iteratively refine each other. **(B)** Schematic representation of the third visual cortical pathway (TVP) specialised for the dynamic aspects of social perception, as proposed by Pitcher and Ungerleider (2021). The pathway originates in the primary visual cortex (V1) and dissociates from both the ventral and dorsal pathways by projecting into the posterior portion of the superior temporal sulcus (pSTS) via motion-selective areas (MT). **(C)** Schematic representation of the multimodal lateral processing pathway (LPP) implementing the sensory processing gateway toward successful face-to-face communication, as proposed in the present perspective. The LPP originates in early visual (V1) and auditory (A1) areas and dissociates from ventral and dorsal pathways by projecting to the mid-posterior portion of STS via animacy and motion-selective areas (here, MT and PT only are represented for clarity of visualisation). Black arrows indicate bottom-up processing along the LPP hierarchy from upstream regions (dashed contours) responsible for multiplex signals (panel A1) to portions of the pSTS (solid contours) that contribute toward the implementation of multimodal gestalts (panel A2). Red arrows indicate top-down multilevel predictions via pSTS to upstream visual and auditory areas (in cyan and yellow, respectively). Panel **(A)** is based on Holler and Levinson, 2019.

information across the entire visual field (Pitcher and Ungerleider, 2021).

Relevantly, Pitcher and Ungerleider (2021) note that the *"proximity (to pSTS, a.n.) of brain areas computing multisensory information relevant to social interactions further dissociates the third pathway from the established role of the ventral and dorsal pathways."* We further elaborate on this by reconceiving the third visual pathway as a fundamental part of a larger multimodal neural system that implements fast analysis of multisensory communicative signals during face-to-face interactions. This pathway projects from early visual and auditory regions along the lateral brain surface and into the pSTS (lateral processing pathway; LPP). From this perspective, regions in the mid-posterior and lateral superior temporal gyrus, which are sensitive to auditory motion, animacy, sounds of moving bodies and dynamic aspects of human vocalisation (i.e., prosodic intonation), become candidate nodes of the auditory bank of LPP.

Analogously to the third visual pathway, evidence supporting the existence of a third lateral auditory cortical pathway, independent of dorsal/ventral pathways (Rauschecker, 1998; Rauschecker and Tian, 2000) and projecting *via* motion-sensitive regions into the posterior STS, comes from both tracer studies in macaques and *in vivo* white matter tractography in humans (see **Table 1**, connectivity profiles). These mid-posterior lateral areas showing anatomical connectivity with the pSTS also show motion-sensitive and voice-sensitive responses, suggesting functional selectivity for dynamic biologically-relevant information along this lateral auditory pathway (see **Table 1**, functional properties). Relevantly, functional interactions and direct anatomical connections have also been observed between

TABLE 1 Functional properties and structural connectivity profile of mid-posterior and lateral auditory areas in the superior temporal gyrus as described in (a) non-human and (b) human primates.

Auditory area	Functional/Connectivity profile	References
(a) In non-human primates		
Mid-posterior parabelt	Auditory motion processing	Poirier et al., 2017
Mid-lateral parabelt	Processing of conspecific vocalization	Petkov et al., 2008; Perrodin et al., 2011
Mid-posterior parabelt	Connection to the mid-posterior STS	Galaburda and Pandya, 1983; Hackett et al., 1998; de la Mothe et al., 2006; Hackett et al., 2007; Smiley et al., 2007
Motion-sensitive areas	Monosynaptic connection to visual MT	Ungerleider and Desimone, 1986; Boussaoud et al., 1990; Palmer and Rosa, 2006
(b) In human primates		
Bilateral hPT	Preferential processing of moving sounds	Krumbholz et al., 2005; Battal et al., 2019
Right lateral hPT	Responses to ipsilateral auditory field	Krumbholz et al., 2005
Bilateral anterior hPT	Encoding of living and human-action sounds categories	Giordano et al., 2013
Right anterior hPT and area adjacent to TVA	Responses to socially meaningful prosody	Belyk and Brown, 2014; Sammler et al., 2015; Hellbernd and Sammler, 2018
Bilateral lateral hPT	White matter connections to mid- and posterior upper bank of STS	Beer et al., 2013
Bilateral mid-lateral STG	White matter connections to mid-upper bank of STS	Beer et al., 2013
Bil. motion-selective portions of hPT	White matter connections to motion-selective hMT	Gurtubay-Antolin et al., 2021

STS, superior temporal sulcus; MT, middle temporal visual area; hPT, human planum temporale; TVA, temporal voice area; STG, superior temporal gyrus; Bil., Bilateral.

auditory and visual motion-sensitive regions (see **Table 1**), suggesting a structural scaffolding for early convergence of multimodal information (Benetti and Collignon, 2022) within temporo-occipital regions of the LPP that might share the same computational goal: fast and reliable analysis of multimodal information relevant to social interactions.

Toward a neurocognitive model of face-to-face communication

In the following section, we attempt to reconcile the psycholinguistic (Holler and Levinson, 2019) and sensory neuroscience (Pitcher and Ungerleider, 2021) frameworks, reviewed so far, toward a coherent neurocognitive model of multimodal face-to-face communication. Accordingly, we propose how key computational principles underlying the perception of multimodal social actions (multiplex signals, multimodal gestalts, and multilevel predictions) might be implemented along the LPP (**Figure 1C**).

Detecting multimodal co-occurrences: Multiplex signals *via* upstream sensory regions

Traditionally, it was thought that multimodal integration takes place in higher-order polysensory areas such as parietal or prefrontal cortices, after unimodal processing in early sensory regions (Ungerleider and Mishkin, 1982; Rauschecker and Tian, 2000); however, accumulating evidence over the past two decades shows clear cross-modal interactions between early sensory areas (Foxe and Schroeder, 2005; Ghazanfar and Schroeder, 2006; Kayser and Logothetis, 2007; Driver and Noesselt, 2008). In fact, several studies with humans (Foxe et al., 2000, 2002; Schürmann et al., 2006; Martuzzi et al., 2007; Besle et al., 2008; Lewis and Noppeney, 2010) and primates (Schroeder et al., 2001; Fu et al., 2003; Kayser et al., 2005, 2008; Lakatos et al., 2007) have proved driving or modulatory effects of cross-modal stimuli at the bottom of the sensory processing hierarchy. Beyond identifying multimodal interactions, such evidence also revealed their ubiquity across the (sub)cortical hierarchy and called for the need to further characterise the computational principles, neural properties and behavioural relevance of these interactions. One possibility is that they differ at different processing stages (i.e., *multistage integration*) along the (sub)cortical hierarchy (Calvert and Thesen, 2004; Noppeney et al., 2018; Noppeney, 2021).

Since visual bodily signals typically precede speech during natural face-to-face interactions (Nota et al., 2021), they may modulate the sound-induced activity in the auditory cortex by resetting the phase of ongoing oscillations (Biau et al., 2015; Mégevand et al., 2020; Pouw et al., 2021). In support of a temporally-sensitive mechanism, neurophysiological (Kayser et al., 2010; Atilgan et al., 2018), and fMRI studies (Lewis and Noppeney, 2010; Werner and Noppeney, 2011) have shown that audiovisual interactions in early auditory cortex and hPT depended on audiovisual temporal coincidence or coherence over time. Sensitivity to temporal co-occurrences is crucial to multiplex signals, which rest on temporal statistical regularities across sensory channels at a perceptual, pre-semantic level (Holler and Levinson, 2019). Therefore, it seems plausible that upstream sensory regions (e.g., visual and auditory cortices) interact in a temporally-sensitive fashion at corresponding processing stages (i.e., *via* multistage integration) to implement multiplex signals [see also Bizley et al. (2016)]. Specifically, it may be that primary visual and auditory cortices concur to support the automatic, salience-driven detection of multimodal co-occurrences, while secondary visual and auditory cortices along the LPP (hMT/EBA and hPT/TVA) concur to represent dynamic aspects of audiovisual bodily signals, mirroring results outside the realm of face-to-face communication (Lewis and Noppeney, 2010).

Recognizing communicative meanings: Multimodal gestalts *via* pSTS

As reviewed above, upstream visual and auditory sensory regions are structurally and functionally interconnected with pSTS. Crucially, this region represents a site of multimodal integration of social and non-social sensory information, as shown in neuroimaging and neurophysiological studies with humans (Beauchamp, 2005; Beauchamp et al., 2008; Werner and Noppeney, 2010a,b; Hirsch et al., 2018; Noah et al., 2020) and non-human primates (Ghazanfar et al., 2008; Froesel et al., 2021). While these studies employed non-linguistic but meaningful world categories such as animals, manipulable objects, and human actions, pSTS is also involved in the processing of communicative and meaningful audiovisual stimuli such as lip-speech (MacSweeney et al., 2000; Wright, 2003; Macaluso et al., 2004; van Atteveldt et al., 2004; Stevenson and James, 2009; Price, 2012; Venezia et al., 2017) and gesture-speech (Holle et al., 2008, 2010; Hubbard et al., 2009; Willems et al., 2016). Consistently, multimodal integration in pSTS may allow the creation of meaningful neural representations (Beauchamp et al., 2004; Noppeney et al., 2018), including those bearing semantic and pragmatic values for social communication (i.e., multimodal gestalts; Holler and Levinson, 2019). In particular, we propose that pSTS might concur toward such (multimodal) neural representations based on Bayesian Causal Inference principles (Körding et al., 2007; Shams and Beierholm, 2010; Noppeney, 2021), mirroring effects found along the dorsal audiovisual pathways for spatial localisation (Rohe and Noppeney, 2015, 2016; Aller and Noppeney, 2019; Ferrari and Noppeney, 2021).

Intriguingly, pSTS is positioned at the intersection of three brain systems respectively responsible for social perception, action observation, and theory of mind (Yang et al., 2015). As noticed by Pitcher and Ungerleider (2021), perceptual analysis of goal-directed actions in the pSTS likely influences activity in parietal and frontal systems that are responsible for action and intention recognition. As such, after receiving converging inputs from upstream sensory regions of the LPP, pSTS may represent the sensory processing gateway that feeds to higher-order networks for social action recognition during face-to-face communication. As a result, multiplex signals may be processed at the semantic and pragmatic levels, enabling the recognition of multimodal gestalts (Holler and Levinson, 2019).

Predicting how the conversation unfolds: Multilevel predictions along the cortical hierarchy

Increasing evidence shows that humans, among other species, build on their past experiences to construct predictive models of themselves and their sensory environment (de Lange et al., 2018). Accordingly, the brain can be conceived as a "prediction machine" (Clark, 2013) that attempts to match bottom-up sensory inputs with top-down expectations. Following hierarchical predictive coding (Rao and Ballard, 1999; Friston, 2005, 2010), any mismatch between expectation and actual input is signalled as a prediction error that propagates up the processing hierarchy to higher-level areas; vice versa, expected inputs are "explained away," resulting in "expectation suppression" (Summerfield et al., 2008; Alink et al., 2010; Richter et al., 2018; Walsh et al., 2020). Importantly, expectation suppression reflects the neural tuning properties along a given processing hierarchy. For example, predictions about visual object and face identity are associated with expectation suppression respectively in object-selective regions (Meyer and Olson, 2011; Kaposvari et al., 2018; Richter et al., 2018; Ferrari et al., 2022; He et al., 2022) and face-selective regions (Summerfield et al., 2008; Amado et al., 2016; Schwiedrzik and Freiwald, 2017) along the ventral visual stream [for corresponding effects in the auditory domain, see e.g., Jaramillo and Zador (2011), Todorovic et al. (2011), Barascud et al. (2016), Heilbron and Chait (2018)].

Similarly, multilevel predictions during face-to-face interactions (Holler and Levinson, 2019) may be implemented *via* mechanisms of hierarchical predictive processing in neural pathways that are responsible for coding the relevant sensory information (e.g., vocal sounds, bodily movements) and linguistic information (e.g., words, sentential units, social actions). Increasing evidence shows signatures of hierarchical predictive processing during language comprehension in left-lateralized fronto-temporal regions of the language network (Blank and Davis, 2016; Sohoglu and Davis, 2016; Willems et al., 2016; Schmitt et al., 2021; Heilbron et al., 2022). Accordingly, predictive processing mechanisms may implement multimodal sensory predictions relevant to face-to-face interactions along the cortical hierarchy of the LPP. Initial evidence shows that hMT and pSTS activity is reduced in response to expected than unexpected visual actions (Koster-Hale and Saxe, 2013), such as human movements violating biomechanical predictions (Costantini et al., 2005; Saygin et al., 2012). Further, pSTS activity is reduced in response to actions that fit rather than violate the spatiotemporal structure of the environment (Koster-Hale and Saxe, 2013), such as shifting head and gaze toward rather than away an abrupt warning signal (Pelphrey et al., 2003). Interestingly, there is evidence of a functional dissociation between hMT and pSTS, with only the latter being sensitive to violations of action intentions (Pelphrey et al., 2004). Such dissociation is suggestive of a hierarchy of computations from sensory processing of dynamic inputs in hMT (at the level of multiplex signals) to semantic and pragmatic analysis in pSTS (at the level of multimodal gestalts), which may then be reflected in the respective expectation suppression profiles. Yet, it remains an open question whether and how multimodal (e.g., audiovisual) predictions arising from face-to-face interactions generate neural signatures of hierarchical predictive processing along the entire LPP, down to upstream sensory regions [for complementary evidence, see Lee and Noppeney (2014)]. Further, it is unknown whether and how higher-order expectations from language, action recognition and theory of mind networks may feed-back to pSTS (Yang et al., 2015) and thus travel down the LPP.

Discussion and conclusion

The current proposal leaves many aspects of the model un- or under-specified, including issues of hemispheric lateralization (Pitcher and Ungerleider, 2021) and the exact relationship between LPP and brain networks responsible for language (Hickok and Poeppel, 2000, 2007; Friederici, 2012; Hagoort, 2019), action recognition (Lingnau and Downing, 2015; Wurm and Caramazza, 2022), and theory of mind (Frith and Frith, 2006; Mar, 2011; Schaafsma et al., 2015). Future research must provide direct empirical evidence to support our framework, as well as refine and enrich it at the algorithmic and neural levels. To start, neuroimaging and neurostimulation techniques may characterise the functional and representational properties of the LPP as proposed here, as well as its degree of lateralization and interconnection with other brain networks (Thiebaut de Schotten and Forkel, 2022). Further, it will be crucial to combine these techniques with methodological approaches that enable human motion-tracking and near-to-optimal preservation of naturalistic, ecological contexts of face-to-face social interactions, such as virtual reality (Peeters, 2019). Complementarily, hyperscanning (Redcay and Schilbach, 2019; Hamilton, 2021) and multibrain stimulation techniques (Novembre and Iannetti, 2021) will be necessary to probe the functional relevance of the LPP during multimodal face-to-face processing across interacting brains. In parallel, the use of computational models (e.g., Bayesian Causal Inference) and neuroscientific-inspired artificial intelligence (i.e., convolutional or deep neural networks) could formalise the empirical evidence and test its role (e.g., necessity, sufficiency) for human behaviour (Hassabis et al., 2017) during face-to-face interactions. Last, but not least, it will be crucial to further embrace an interdisciplinary perspective in which psycholinguistics and neuroscientific frameworks would be reciprocally validated.

We conclude that the time is mature to accept the challenge we, among others before, advocated in this perspective and move beyond the speech-centred perspective dominating research on the neurocognitive mechanisms of human communication and language. We offer an original perspective bridging two recent propositions in psycholinguistics (Holler and Levinson, 2019) and sensory neuroscience (Pitcher and Ungerleider, 2021) into a neurocognitive model of multimodal face-to-face communication. Testing this framework represents a novel and promising endeavour for future research.

Author contributions

SB and AF contributed equally to the original conception of the perspective and wrote the first draft of the manuscript. FP contributed to further developing the preliminary conception. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

SB was supported by a "Starting Grant DM 737/21" from the University of Trento (R06). SB and FP were supported by a "Progetto di Rilevante Interesse Nazionale (PRIN)" from the Italian Ministry for Education, University and Research (MIUR-PRIN 2017 n.20177894ZH).

Acknowledgments

We would like to express our gratitude to Eugenio Parise for providing insightful comments on the preliminary version of this perspective.

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.

The reviewer JT declared a shared parent affiliation with the author AF, and the handling editor declared a past collaboration with the author SB, at the time of review.

Publisher's note

References

Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., and Muckli, L. (2010). Stimulus predictability reduces responses in primary visual cortex. *J. Neurosci.* 30, 2960–2966. doi: 10.1523/JNEUROSCI.3730-10.2010

Aller, M., and Noppeney, U. (2019). To integrate or not to integrate: Temporal dynamics of hierarchical Bayesian causal inference. *PLoS Biol.* 17:e3000210. doi: 10.1371/journal.pbio.3000210

Allison, T., Puce, A., and McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends Cogn. Sci.* 4, 267–278. doi: 10.1016/S1364-6613(00)01501-1

Amado, C., Hermann, P., Kovács, P., Grotheer, M., Vidnyánszky, Z., and Kovács, G. (2016). The contribution of surprise to the prediction based modulation of fMRI responses. *Neuropsychologia* 84, 105–112. doi: 10.1016/j.neuropsychologia.2016.02.003

Atilgan, H., Town, S. M., Wood, K. C., Jones, G. P., Maddox, R. K., Lee, A. K. C., et al. (2018). Integration of visual information in auditory cortex promotes auditory scene analysis through multisensory binding. *Neuron* 97, 640–655.e4. doi: 10.1016/j.neuron.2017.12.034

Barascud, N., Pearce, M. T., Griffiths, T. D., Friston, K. J., and Chait, M. (2016). Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. *Proc. Natl. Acad. Sci. U.S.A.* 113:E616–E625. doi: 10.1073/pnas.1508523113

Battal, C., Rezk, M., Mattioni, S., Vadlamudi, J., and Collignon, O. (2019). Representation of auditory motion directions and sound source locations in the human planum temporale. *J. Neurosci.* 39, 2208–2220. doi: 10.1523/JNEUROSCI.2289-18.2018

Beauchamp, M. S. (2005). See me, hear me, touch me: Multisensory integration in lateral occipital-temporal cortex. *Curr. Opin. Neurobiol.* 15, 145–153. doi: 10.1016/j.conb.2005.03.011

Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., and Martin, A. (2004). Unraveling multisensory integration: Patchy organization within human STS multisensory cortex. *Nat. Neurosci.* 7, 1190–1192. doi: 10.1038/nn1333

Beauchamp, M. S., Yasar, N. E., Frye, R. E., and Ro, T. (2008). Touch, sound and vision in human superior temporal sulcus. *Neuroimage* 41, 1011–1020. doi: 10.1016/j.neuroimage.2008.03.015

Beer, A. L., Plank, T., Meyer, G., and Greenlee, M. W. (2013). Combined diffusion-weighted and functional magnetic resonance imaging reveals a temporal-occipital network involved in auditory-visual object processing. *Front. Integr. Neurosci.* 7:5. doi: 10.3389/fnint.2013.00005

Belyk, M., and Brown, S. (2014). Perception of affective and linguistic prosody: An ALE meta-analysis of neuroimaging studies. *Soc. Cogn. Affect. Neurosci.* 9, 1395–1403. doi: 10.1093/scan/nst124

Benetti, S., and Collignon, O. (2022). Cross-modal integration and plasticity in the superior temporal cortex. *Handb. Clin. Neurol.* 187, 127–143. doi: 10.1016/B978-0-12-823493-8.00026-2

Besle, J., Fischer, C., Bidet-Caulet, A., Lecaignard, F., Bertrand, O., and Giard, M.-H. (2008). Visual activation and audiovisual interactions in the auditory cortex during speech perception: Intracranial recordings in humans. *J. Neurosci.* 28, 14301–14310. doi: 10.1523/JNEUROSCI.2875-08.2008

Biau, E., Torralba, M., Fuentemilla, L., de Diego Balaguer, R., and Soto-Faraco, S. (2015). Speaker's hand gestures modulate speech perception through phase resetting of ongoing neural oscillations. *Cortex* 68, 76–85. doi: 10.1016/j.cortex.2014.11.018

Bizley, J. K., Maddox, R. K., and Lee, A. K. C. (2016). Defining auditory-visual objects: Behavioral tests and physiological mechanisms. *Trends Neurosci.* 39, 74–85. doi: 10.1016/j.tins.2015.12.007

Blank, H., and Davis, M. H. (2016). Prediction Errors but Not Sharpened Signals Simulate Multivoxel fMRI Patterns during Speech Perception. *PLoS Biol.* 14:e1002577. doi: 10.1371/journal.pbio.1002577

Boussaoud, D., Ungerleider, L. G., and Desimone, R. (1990). Pathways for motion analysis: Cortical connections of the medial superior temporal and fundus of the superior temporal visual areas in the macaque. *J. Comp. Neurol.* 296, 462–495. doi: 10.1002/CNE.902960311

Bruce, C., Desimone, R., and Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J. Neurophysiol.* 46, 369–384. doi: 10.1152/jn.1981.46.2.369

Calvert, G. A., and Thesen, T. (2004). Multisensory integration: Methodological approaches and emerging principles in the human brain. *J. Physiol. Paris* 98, 191–205. doi: 10.1016/j.jphysparis.2004.03.018

Cameron-Faulkner, T., Theakston, A., Lieven, E., and Tomasello, M. (2015). The relationship between infant holdout and gives, and pointing. *Infancy* 20, 576–586. doi: 10.1111/infa.12085

Chen, L., and Vroomen, J. (2013). Intersensory binding across space and time: A tutorial review. *Atten. Percept. Psychophys.* 75, 790–811. doi: 10.3758/s13414-013-0475-4

Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477

Costantini, M., Galati, G., Ferretti, A., Caulo, M., Tartaro, A., Romani, G. L., et al. (2005). Neural systems underlying observation of humanly impossible movements: An FMRI study. *Cereb. Cortex* 15, 1761–1767. doi: 10.1093/cercor/bhi053

de la Mothe, L. A., Blumell, S., Kajikawa, Y., and Hackett, T. A. (2006). Cortical connections of the auditory cortex in marmoset monkeys: Core and medial belt regions. *J. Comp. Neurol.* 496, 27–71. doi: 10.1002/cne.20923

de Lange, F. P., Heilbron, M., and Kok, P. (2018). How do expectations shape perception? *Trends Cogn. Sci.* 22, 764–779. doi: 10.1016/j.tics.2018.06.002

Deen, B., Saxe, R., and Kanwisher, N. (2020). Processing communicative facial and vocal cues in the superior temporal sulcus. *Neuroimage* 221:117191. doi: 10.1016/J.NEUROIMAGE.2020.117191

Desimone, R., and Ungerleider, L. G. (1986). Multiple visual areas in the caudal superior temporal sulcus of the macaque. *J. Comp. Neurol.* 248, 164–189. doi: 10.1002/cne.902480203

Diederich, N. J., Fénelon, G., Stebbins, G., and Goetz, C. G. (2009). Hallucinations in Parkinson disease. *Nat. Rev. Neurol.* 5, 331–342.

Drijvers, L., and Holler, J. (2022). The multimodal facilitation effect in human communication. *Psychon. Bull. Rev.* [Epub ahead of print]. doi: 10.3758/S13423-022-02178-X

Driver, J., and Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on "sensory-specific" brain regions, neural responses, and judgments. *Neuron* 57, 11–23. doi: 10.1016/j.neuron.2007.12.013

Ferrari, A., and Noppeney, U. (2021). Attention controls multisensory perception via two distinct mechanisms at different levels of the cortical hierarchy. *PLoS Biol.* 19:e3001465. doi: 10.1371/journal.pbio.3001465

Ferrari, A., Richter, D., and de Lange, F. P. (2022). Updating contextual sensory expectations for adaptive behaviour. *J. Neurosci.* 42, 8855–8869. doi: 10.1523/JNEUROSCI.1107-22.2022

Finzi, D., Gomez, J., Nordt, M., Rezai, A. A., Poltoratski, S., and Grill-Spector, K. (2021). Differential spatial computations in ventral and lateral face-selective regions are scaffolded by structural connections. *Nat. Commun.* 12:2278. doi: 10.1038/s41467-021-22524-2

Foxe, J. J., Morocz, I. A., Murray, M. M., Higgins, B. A., Javitt, D. C., and Schroeder, C. E. (2000). Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Cogn. Brain Res.* 10, 77–83. doi: 10.1016/S0926-6410(00)00024-0

Foxe, J. J., and Schroeder, C. E. (2005). The case for feedforward multisensory convergence during early cortical processing. *Neuroreport* 16, 419–423. doi: 10.1097/00001756-200504040-00001

Foxe, J. J., Wylie, G. R., Martinez, A., Schroeder, C. E., Javitt, D. C., Guilfoyle, D., et al. (2002). Auditory-somatosensory multisensory processing in auditory association cortex: An fMRI study. *J. Neurophysiol.* 88, 540–543. doi: 10.1152/jn.00694.2001

Friederici, A. D. (2012). The cortical language circuit: From auditory perception to sentence comprehension. *Trends Cogn. Sci.* 16, 262–268. doi: 10.1016/j.tics.2012.04.001

Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. B* 360, 815–836. doi: 10.1098/rstb.2005.1622

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Frith, C. D., and Frith, U. (2006). The neural basis of mentalizing. *Neuron* 50, 531–534. doi: 10.1016/j.neuron.2006.05.001

Froesel, M., Gacoin, M., Clavagnier, S., Hauser, M., Goudard, Q., and Hamed, S. (2021). Neural correlates of audio-visual integration of socially meaningful information in macaque monkeys. *bioRxiv* [Preprint]. doi: 10.1101/2021.05.02.442333

Fu, K. M. G., Johnston, T. A., Shah, A. S., Arnold, L., Smiley, J., Hackett, T. A., et al. (2003). Auditory cortical neurons respond to somatosensory stimulation. *J. Neurosci.* 23, 7510–7515.

Galaburda, A. M., and Pandya, D. N. (1983). The intrinsic architectonic and connectional organization of the superior temporal region of the rhesus monkey. *J. Comp. Neurol.* 221, 169–184. doi: 10.1002/CNE.902210206

Gallagher, S. (ed.) (2020). "Direct Social Perception," in *Action and Interaction*, (Oxford: Oxford University Press), 121–154. doi: 10.1093/oso/9780198846345.003.0007

Ghazanfar, A., and Schroeder, C. (2006). Is neocortex essentially multisensory? *Trends Cogn. Sci.* 10, 278–285. doi: 10.1016/j.tics.2006.04.008

Ghazanfar, A. A., Chandrasekaran, C., and Logothetis, N. K. (2008). Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys. *J. Neurosci.* 28, 4457–4469. doi: 10.1523/JNEUROSCI.0541-08.2008

Gingras, G., Rowland, B. A., and Stein, B. E. (2009). The differing impact of multisensory and unisensory integration on behavior. *J. Neurosci.* 29, 4897–4902. doi: 10.1523/JNEUROSCI.4120-08.2009

Giordano, B. L., McAdams, S., Zatorre, R. J., Kriegeskorte, N., and Belin, P. (2013). Abstract encoding of auditory objects in cortical activity patterns. *Cereb. Cortex* 23, 2025–2037. doi: 10.1093/CERCOR/BHS162

Gurtubay-Antolin, A., Battal, C., Maffei, C., Rezk, M., Mattioni, S., Jovicich, J., et al. (2021). Direct structural connections between auditory and visual motion-selective regions in humans. *J. Neurosci.* 41, 2393–2405. doi: 10.1523/jneurosci.1552-20.2021

Hackett, T. A., Smiley, J. F., Ulbert, I., Karmos, G., Lakatos, P., De La Mothe, L. A., et al. (2007). Sources of somatosensory input to the caudal belt areas of auditory cortex. *Perception* 36, 1419–1430. doi: 10.1068/p5841

Hackett, T. A., Stepniewska, I., and Kaas, J. H. (1998). Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J. Comp. Neurol.* 394, 475-495. doi: 10.1002/(SICI)1096-9861(19980518)394:4

Hagoort, P. (2005). On broca, brain, and binding: A new framework. *Trends Cogn. Sci.* 9, 416–423. doi: 10.1016/j.tics.2005.07.004

Hagoort, P. (2019). The neurobiology of language beyond single-word processing. *Science* 366, 55–58. doi: 10.1126/science.aax0289

Hamilton, A. F. C. (2021). Hyperscanning: Beyond the Hype. *Neuron* 109, 404–407. doi: 10.1016/j.neuron.2020.11.008

Hassabis, D., Kumaran, D., Summerfield, C., and Botvinick, M. (2017). Neuroscience-Inspired Artificial Intelligence. *Neuron* 95, 245–258. doi: 10.1016/j.neuron.2017.06.011

Hasson, U., Egidi, G., Marelli, M., and Willems, R. M. (2018). Grounding the neurobiology of language in first principles: The necessity of non-language-centric explanations for language comprehension. *Cognition* 180, 135–157. doi: 10.1016/J.COGNITION.2018.06.018

He, T., Richter, D., Wang, Z., and de Lange, F. P. (2022). Spatial and temporal context jointly modulate the sensory response within the ventral visual stream. *J. Cogn. Neurosci.* 34, 332–347. doi: 10.1162/jocn_a_01792

Heilbron, M., Armeni, K., Schoffelen, J. M., Hagoort, P., and de Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. *Proc. Natl. Acad. Sci. U.S.A.* 119:e2201968119. doi: 10.1073/pnas.2201968119

Heilbron, M., and Chait, M. (2018). Great expectations: Is there evidence for predictive coding in auditory cortex? *Neuroscience* 389, 54–73. doi: 10.1016/j.neuroscience.2017.07.061

Hein, G., and Knight, R. T. (2008). Superior temporal sulcus - it's my area: Or is it? *J. Cogn. Neurosci.* 20, 2125–2136. doi: 10.1162/jocn.2008.20148

Hellbernd, N., and Sammler, D. (2018). Neural bases of social communicative intentions in speech. *Soc. Cogn. Affect. Neurosci.* 13, 604–615. doi: 10.1093/SCAN/NSY034

Hickok, G., and Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.* 4, 131–138. doi: 10.1016/S1364-6613(00)01463-7

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113

Hirsch, J., Noah, J. A., Zhang, X., Dravida, S., and Ono, Y. (2018). A cross-brain neural mechanism for human-to-human verbal communication. *Soc. Cogn. Affect. Neurosci.* 13, 907–920. doi: 10.1093/scan/nsy070

Holle, H., Gunter, T. C., Rüschemeyer, S. A., Hennenlotter, A., and Iacoboni, M. (2008). Neural correlates of the processing of co-speech gestures. *Neuroimage* 39, 2010–2024. doi: 10.1016/j.neuroimage.2007.10.055

Holle, H., Obleser, J., Rueschemeyer, S. A., and Gunter, T. C. (2010). Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *Neuroimage* 49, 875–884. doi: 10.1016/j.neuroimage.2009.08.058

Holler, J. (2022). Visual bodily signals as core devices for coordinating minds in interaction. *Philos. Trans. R. Soc. B* 377:20210094. doi: 10.1098/rstb.2021.0094

Holler, J., Kendrick, K. H., and Levinson, S. C. (2018). Processing language in face-to-face conversation: Questions with gestures get faster responses. *Psychon. Bull. Rev.* 25, 1900–1908. doi: 10.3758/s13423-017-1363-z

Holler, J., and Levinson, S. C. (2019). Multimodal Language Processing in Human Communication. *Trends Cogn. Sci.* 23, 639–652. doi: 10.1016/j.tics.2019.05.006

Hubbard, A. L., Wilson, S. M., Callan, D. E., and Dapretto, M. (2009). Giving speech a hand: Gesture modulates activity in auditory cortex during speech perception. *Hum. Brain Mapp.* 30, 1028–1037. doi: 10.1002/hbm.20565

Huk, A. C., Dougherty, R. F., and Heeger, D. J. (2002). Retinotopy and functional subdivision of human areas MT and MST. *J. Neurosci.* 22, 7195–7205. doi: 10.1523/jneurosci.22-16-07195.2002

Jaramillo, S., and Zador, A. M. (2011). The auditory cortex mediates the perceptual effects of acoustic temporal expectation. *Nat. Neurosci.* 14, 246–253. doi: 10.1038/nn.2688

Kaposvari, P., Kumar, S., and Vogels, R. (2018). Statistical learning signals in macaque inferior temporal cortex. *Cereb. Cortex* 28, 250–266. doi: 10.1093/cercor/bhw374

Kayser, C., and Logothetis, N. K. (2007). Do early sensory cortices integrate cross-modal information? *Brain Struct. Funct.* 212, 121–132. doi: 10.1007/s00429-007-0154-0

Kayser, C., Logothetis, N. K., and Panzeri, S. (2010). Visual enhancement of the information representation in auditory cortex. *Curr. Biol.* 20, 19–24. doi: 10.1016/j.cub.2009.10.068

Kayser, C., Petkov, C. I., Augath, M., and Logothetis, N. K. (2005). Integration of touch and sound in auditory cortex. *Neuron* 48, 373–384. doi: 10.1016/j.neuron.2005.09.018

Kayser, C., Petkov, C. I., and Logothetis, N. K. (2008). Visual modulation of neurons in auditory cortex. *Cereb. Cortex* 18, 1560–1574. doi: 10.1093/cercor/bhm187

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., and Shams, L. (2007). Causal inference in multisensory perception. *PLoS One* 2:e943. doi: 10.1371/journal.pone.0000943

Koster-Hale, J., and Saxe, R. (2013). Theory of mind: A neural prediction problem. *Neuron* 79, 836–848. doi: 10.1016/j.neuron.2013.08.020

Krumbholz, K., Schönwiesner, M., Von Cramon, D. Y., Rübsamen, R., Shah, N. J., Zilles, K., et al. (2005). Representation of interaural temporal information from left and right auditory space in the human planum temporale and inferior parietal lobe. *Cereb. Cortex* 15, 317–324. doi: 10.1093/CERCOR/BHH133

Lakatos, P., Chen, C. M., O'Connell, M. N., Mills, A., and Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* 53, 279–292. doi: 10.1016/j.neuron.2006.12.011

Landsiedel, J., Daughters, K., Downing, P. E., and Koldewyn, K. (2022). The role of motion in the neural representation of social interactions in the posterior temporal cortex. *Neuroimage* 262, 119533. doi: 10.1016/j.neuroimage.2022.119533

Lee, H., and Noppeney, U. (2014). Temporal prediction errors in visual and auditory cortices. *Curr. Biol.* 24:R309–R310. doi: 10.1016/j.cub.2014.02.007

Levinson, S. C. (2016). Turn-taking in human communication – origins and implications for language processing. *Trends Cogn. Sci.* 20, 6–14. doi: 10.1016/J.TICS.2015.10.010

Levinson, S. C., and Holler, J. (2014). The origin of human multi-modal communication. *Philos. Trans. R. Soc. B* 369:20130302. doi: 10.1098/rstb.2013.0302

Lewis, R., and Noppeney, U. (2010). Audiovisual synchrony improves motion discrimination via enhanced connectivity between early visual and auditory areas. *J. Neurosci.* 30, 12329–12339. doi: 10.1523/JNEUROSCI.5745-09.2010

Lingnau, A., and Downing, P. E. (2015). The lateral occipitotemporal cortex in action. *Trends Cogn. Sci.* 19, 268–277. doi: 10.1016/j.tics.2015.03.006

Macaluso, E., George, N., Dolan, R., Spence, C., and Driver, J. (2004). Spatial and temporal factors during processing of audiovisual speech: A PET study. *Neuroimage* 21, 725–732. doi: 10.1016/j.neuroimage.2003.09.049

MacSweeney, M., Amaro, E., Calvert, G. A., Campbell, R., David, A. S., McGuire, P., et al. (2000). Silent speechreading in the absence of scanner noise: An event-related fMRI study. *Neuroreport* 11:1729. doi: 10.1097/00001756-200006050-00026

Mar, R. A. (2011). The neural bases of social cognition and story comprehension. *Annu. Rev. Psychol.* 62, 103–134. doi: 10.1146/annurev-psych-120709-145406

Martuzzi, R., Murray, M. M., Michel, C. M., Thiran, J. P., Maeder, P. P., Clarke, S., et al. (2007). Multisensory interactions within human primary cortices revealed by BOLD dynamics. *Cereb. Cortex* 17, 1672–1679. doi: 10.1093/cercor/bhl077

Mégevand, P., Mercier, M. R., Groppe, D. M., Golumbic, E. Z., Mesgarani, N., Beauchamp, M. S., et al. (2020). Crossmodal phase reset and evoked responses provide complementary mechanisms for the influence of visual speech in auditory cortex. *J. Neurosci.* 40, 8530–8542. doi: 10.1523/JNEUROSCI.0555-20.2020

Meyer, T., and Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 108, 19401–19406. doi: 10.1073/pnas.1112895108

Murray, M. M., Foxe, J. J., Higgins, B. A., Javitt, D. C., and Schroeder, C. E. (2001). Visuo-spatial neural response interactions in early cortical processing during a simple reaction time task: A high-density electrical mapping study. *Neuropsychologia* 39, 828–844. doi: 10.1016/S0028-3932(01)00004-5

Noah, J. A., Zhang, X., Dravida, S., Ono, Y., Naples, A., McPartland, J. C., et al. (2020). Real-time eye-to-eye contact is associated with cross-brain neural coupling in angular gyrus. *Front. Hum. Neurosci.* 14:19. doi: 10.3389/fnhum.2020.00019

Noppeney, U. (2021). Perceptual inference, learning, and attention in a multisensory world. *Annu. Rev. Neurosci.* 44, 449–473. doi: 10.1146/annurev-neuro-100120-085519

Noppeney, U., Jones, S. A., Rohe, T., and Ferrari, A. (2018). See what you hear-How the brain forms representations across the senses. *Neuroforum* 24, 237–246. doi: 10.1515/nf-2017-A066

Nota, N., Trujillo, J. P., and Holler, J. (2021). Facial signals and social actions in multimodal face-to-face interaction. *Brain Sci.* 11:1017. doi: 10.3390/brainsci11081017

Novembre, G., and Iannetti, G. D. (2021). Hyperscanning alone cannot prove causality. Multibrain stimulation can. *Trends Cogn. Sci.* 25, 96–99. doi: 10.1016/j.tics.2020.11.003

Özyürek, A. (2014). Hearing and seeing meaning in speech and gesture: Insights from brain and behaviour. *Philos. Trans. R. Soc. B* 369:20130296. doi: 10.1098/rstb.2013.0296

Palmer, S. M., and Rosa, M. G. P. (2006). Quantitative analysis of the corticocortical projections to the middle temporal area in the marmoset monkey: Evolutionary and functional implications. *Cereb. Cortex* 16, 1361–1375. doi: 10.1093/cercor/bhj078

Peeters, D. (2019). Virtual reality: A game-changing method for the language sciences. *Psychon. Bull. Rev.* 26, 894–900. doi: 10.3758/s13423-019-01571-3

Pelphrey, K. A., Mitchell, T. V., McKeown, M. J., Goldstein, J., Allison, T., and McCarthy, G. (2003). Brain activity evoked by the perception of human walking: Controlling for meaningful coherent motion. *J. Neurosci.* 23, 6819–6825. doi: 10.1523/JNEUROSCI.23-17-06819.2003

Pelphrey, K. A., Morris, J. P., and McCarthy, G. (2004). Grasping the intentions of others: The perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *J. Cogn. Neurosci.* 16, 1706–1716. doi: 10.1162/0898929042947900

Perniss, P. (2018). Why we should study multimodal language. *Front. Psychol.* 9:1109. doi: 10.3389/FPSYG.2018.01109/BIBTEX

Perrodin, C., Kayser, C., Logothetis, N. K., and Petkov, C. I. (2011). Voice cells in the primate temporal lobe. *Curr. Biol.* 21, 1408–1415. doi: 10.1016/j.cub.2011.07.028

Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., and Logothetis, N. K. (2008). A voice region in the monkey brain. *Nat. Neurosci.* 11, 367–374.

Pitcher, D., Duchaine, B., and Walsh, V. (2014). Combined TMS and fMRI reveal dissociable cortical pathways for dynamic and static face perception. *Curr. Biol.* 24, 2066–2070. doi: 10.1016/j.cub.2014.07.060

Pitcher, D., Pilkington, A., Rauth, L., Baker, C., Kravitz, D. J., and Ungerleider, L. G. (2020). The human posterior superior temporal sulcus samples visual space differently from other face-selective regions. *Cereb. Cortex* 30, 778–785. doi: 10.1093/cercor/bhz125

Pitcher, D., and Ungerleider, L. G. (2021). Evidence for a third visual pathway specialized for social perception. *Trends Cogn. Sci.* 25, 100–110. doi: 10.1016/j.tics.2020.11.006

Poirier, C., Baumann, S., Dheerendra, P., Joly, O., Hunter, D., Balezeau, F., et al. (2017). Auditory motion-specific mechanisms in the primate brain. *PLoS Biol.* 15:e2001379. doi: 10.1371/JOURNAL.PBIO.2001379

Pouw, W., Proksch, S., Drijvers, L., Gamba, M., Holler, J., Kello, C., et al. (2021). Multilevel rhythms in multimodal communication. *Philos. Trans. R. Soc. B* 376: 20200334. doi: 10.31219/OSF.IO/PSMHN

Price, C. J. (2012). A review and synthesis of the first 20years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* 62, 816–847. doi: 10.1016/j.neuroimage.2012.04.062

Rao, R. P. N., and Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580

Rauschecker, J. P. (1998). Cortical processing of complex sounds. *Curr. Opin. Neurobiol.* 8, 516–521. doi: 10.1016/S0959-4388(98)80040-8

Rauschecker, J. P., and Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11800–11806. doi: 10.1073/pnas.97.22.11800

Redcay, E., and Carlson, T. A. (2015). Rapid neural discrimination of communicative gestures. *Soc. Cogn. Affect. Neurosci.* 10, 545–551. doi: 10.1093/scan/nsu089

Redcay, E., and Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nat. Rev. Neurosci.* 20, 495–505. doi: 10.1038/s41583-019-0179-4

Richter, D., Ekman, M., and de Lange, F. P. (2018). Suppressed sensory response to predictable object stimuli throughout the ventral visual stream. *J. Neurosci.* 38, 7452–7461. doi: 10.1523/JNEUROSCI.3421-17.2018

Rohe, T., and Noppeney, U. (2015). Cortical hierarchies perform bayesian causal inference in multisensory perception. *PLoS Biol.* 13:e1002073. doi: 10.1371/journal.pbio.1002073

Rohe, T., and Noppeney, U. (2016). Distinct computational principles govern multisensory integration in primary sensory and association cortices. *Curr. Biol.* 26, 509–514. doi: 10.1016/j.cub.2015.12.056

Rossano, F., and Liebal, K. (2014). "Requests and offers," in orangutans and human infants," in *Requesting in social interaction*, ed. P. Drew (Amsterdam: Benjamins), 335–364. doi: 10.1075/slsi.26.13ros

Sammler, D., Grosbras, M. H., Anwander, A., Bestelmeyer, P. E. G., and Belin, P. (2015). Dorsal and ventral pathways for prosody. *Curr. Biol.* 25, 3079–3085. doi: 10.1016/J.CUB.2015.10.009

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., and Frith, C. (2012). The thing that should not be: Predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Soc. Cogn. Affect. Neurosci.* 7, 413–422. doi: 10.1093/scan/nsr025

Schaafsma, S. M., Pfaff, D. W., Spunt, R. P., and Adolphs, R. (2015). Deconstructing and reconstructing theory of mind. *Trends Cogn. Sci.* 19, 65–72. doi: 10.1016/j.tics.2014.11.007

Schmitt, L. M., Erb, J., Tune, S., Rysop, A. U., Hartwigsen, G., and Obleser, J. (2021). Predicting speech from a cortical hierarchy of event-based time scales. *Sci. Adv.* 7:eabi6070. doi: 10.1126/sciadv.abi6070

Schroeder, C. E., Lindsley, R. W., Specht, C., Marcovici, A., Smiley, J. F., and Javitt, D. C. (2001). Somatosensory input to auditory association cortex in the macaque monkey. *J. Neurophysiol.* 85, 1322–1327.

Schürmann, M., Caetano, G., Hlushchuk, Y., Jousmäki, V., and Hari, R. (2006). Touch activates human auditory cortex. *Neuroimage* 30, 1325–1331. doi: 10.1016/j.neuroimage.2005.11.020

Schwiedrzik, C. M., and Freiwald, W. A. (2017). High-level prediction signals in a low-level area of the macaque face-processing hierarchy. *Neuron* 96, 89–97.e4. doi: 10.1016/j.neuron.2017.09.007

Senkowski, D. (2005). Oscillatory beta activity predicts response speed during a multisensory audiovisual reaction time task: A high-density electrical mapping study. *Cereb. Cortex* 16, 1556–1565. doi: 10.1093/cercor/bhj091

Shams, L., and Beierholm, U. R. (2010). Causal inference in perception. *Trends Cogn. Sci.* 14, 425–432. doi: 10.1016/j.tics.2010.07.001

Smiley, J. F., Hackett, T. A., Ulbert, I., Karmas, G., Lakatos, P., Javitt, D. C., et al. (2007). Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *J. Comp. Neurol.* 502, 894–923. doi: 10.1002/CNE.21325

Sohoglu, E., and Davis, M. H. (2016). Perceptual learning of degraded speech by minimizing prediction error. *Proc. Natl. Acad. Sci. U.S.A.* 113:E1747–E1756. doi: 10.1073/pnas.1523266113

Sperber, D., and Wilson, D. (1995). *Relevance: Communication and cognition*, 2nd Edn. Hoboken, NJ: Blackwell Publishing.

Stein, B. E. (2012). *The new handbook of multisensory processing*. Cambridge, MA: The MIT Press. doi: 10.7551/mitpress/8466.003.0001

Stevenson, R. A., and James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *Neuroimage* 44, 1210–1223. doi: 10.1016/j.neuroimage.2008.09.034

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., and Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nat. Neurosci.* 11, 1004–1006. doi: 10.1038/nn.2163

Thiebaut de Schotten, M., and Forkel, S. J. (2022). The emergent properties of the connected brain. *Science* 378, 505–510. doi: 10.1126/science.abq2591

Todorovic, A., van Ede, F., Maris, E., and de Lange, F. P. (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: An MEG study. *J. Neurosci.* 31, 9118–9123. doi: 10.1523/JNEUROSCI.1425-11.2011

Tomasello, R., Grisoni, L., Boux, I., Sammler, D., and Pulvermüller, F. (2022). Instantaneous neural processing of communicative functions conveyed by speech prosody. *Cereb. Cortex* 32, 4885–4901. doi: 10.1093/cercor/bhab522

Trujillo, J. P., and Holler, J. (2023). Interactionally embedded gestalt principles of multimodal human communication. *Perspect. Psychol. Sci.* doi: 10.1177/17456916221141422

Ungerleider, L. G., and Desimone, R. (1986). Cortical connections of visual area MT in the macaque. *J. Comp. Neurol.* 248, 190–222. doi: 10.1002/CNE.902480204

Ungerleider, L. G., and Mishkin, M. (1982). "Two cortical visual systems," in *Analysis of visual behavior*, eds D. Ingle, M. Goodale, and R. Mansfield (Cambridge: MIT Press).

van Atteveldt, N., Formisano, E., Goebel, R., and Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron* 43, 271–282. doi: 10.1016/j.neuron.2004.06.025

Venezia, J. H., Vaden, K. I., Rong, F., Maddox, D., Saberi, K., and Hickok, G. (2017). Auditory, visual and audiovisual speech processing streams in superior temporal sulcus. *Front. Hum. Neurosci.* 11:174. doi: 10.3389/fnhum.2017.00174

Vigliocco, G., Perniss, P., and Vinson, D. (2014). Language as a multimodal phenomenon: Implications for language learning, processing and evolution. *Philos. Trans. R. Soc. B* 369:20130292. doi: 10.1098/RSTB.2013.0292

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., et al. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychol. Bull.* 138, 1172–1217. doi: 10.1037/a0029333

Walbrin, J., and Koldewyn, K. (2019). Dyadic interaction processing in the posterior temporal cortex. *Neuroimage* 198, 296–302. doi: 10.1016/j.neuroimage.2019.05.027

Walsh, K. S., McGovern, D. P., Clark, A., and O'Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Ann. N. Y. Acad. Sci.* 1464, 242–268. doi: 10.1111/nyas.14321

Werner, S., and Noppeney, U. (2010b). Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. *Cereb. Cortex* 20, 1829–1842. doi: 10.1093/cercor/bhp248

Werner, S., and Noppeney, U. (2010a). Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *J. Neurosci.* 30, 2662–2675. doi: 10.1523/JNEUROSCI.5091-09.2010

Werner, S., and Noppeney, U. (2011). The contributions of transient and sustained response codes to audiovisual integration. *Cereb. Cortex* 21, 920–931. doi: 10.1093/cercor/bhq161

Willems, R. M., Frank, S. L., Nijhof, A. D., Hagoort, P., and van den Bosch, A. (2016). Prediction during natural language comprehension. *Cereb. Cortex* 26, 2506–2516. doi: 10.1093/cercor/bhv075

Wright, T. M. (2003). Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cereb. Cortex* 13, 1034–1043. doi: 10.1093/cercor/13.10.1034

Wurm, M. F., and Caramazza, A. (2022). Two 'what' pathways for action and object recognition. *Trends Cogn. Sci.* 26, 103–116. doi: 10.1016/j.tics.2021.10.003

Yang, D. Y. J., Rosenblau, G., Keifer, C., and Pelphrey, K. A. (2015). An integrative neural model of social perception, action observation, and theory of mind. *Neurosci. Biobehav. Rev.* 51, 263–275. doi: 10.1016/j.neubiorev.2015.01.020

TYPE Original Research
PUBLISHED 08 February 2023
DOI 10.3389/fnhum.2023.1049615

OPEN ACCESS

EDITED BY
Alessia Tonelli,
Italian Institute of Technology (IIT), Italy

REVIEWED BY
Christopher DiMattina,
Florida Gulf Coast University, United States
Karola Schlegelmilch,
Max Planck Institute for Human Development, Germany

*CORRESPONDENCE
Maria Michela Del Viva
✉ maria.delviva@unifi.it

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

SPECIALTY SECTION
This article was submitted to
Sensory Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 20 September 2022
ACCEPTED 18 January 2023
PUBLISHED 08 February 2023

CITATION
Castellotti S, D'Agostino O and Del Viva MM (2023) Fast discrimination of fragmentary images: the role of local optimal information. *Front. Hum. Neurosci.* 17:1049615.
doi: 10.3389/fnhum.2023.1049615

Fast discrimination of fragmentary images: the role of local optimal information

Serena Castellotti[†], Ottavia D'Agostino[†] and Maria Michela Del Viva*

Department of Neurofarba, University of Florence, Florence, Italy

In naturalistic conditions, objects in the scene may be partly occluded and the visual system has to recognize the whole image based on the little information contained in some visible fragments. Previous studies demonstrated that humans can successfully recognize severely occluded images, but the underlying mechanisms occurring in the early stages of visual processing are still poorly understood. The main objective of this work is to investigate the contribution of local information contained in a few visible fragments to image discrimination in fast vision. It has been already shown that a specific set of features, predicted by a constrained maximum-entropy model to be optimal carriers of information (optimal features), are used to build simplified early visual representations (primal sketch) that are sufficient for fast image discrimination. These features are also considered salient by the visual system and can guide visual attention when presented isolated in artificial stimuli. Here, we explore whether these local features also play a significant role in more natural settings, where all existing features are kept, but the overall available information is drastically reduced. Indeed, the task requires discrimination of naturalistic images based on a very brief presentation (25 ms) of a few small visible image fragments. In the main experiment, we reduced the possibility to perform the task based on global-luminance positional cues by presenting randomly inverted-contrast images, and we measured how much observers' performance relies on the local features contained in the fragments or on global information. The size and the number of fragments were determined in two preliminary experiments. Results show that observers are very skilled in fast image discrimination, even when a drastic occlusion is applied. When observers cannot rely on the position of global-luminance information, the probability of correct discrimination increases when the visible fragments contain a high number of optimal features. These results suggest that such optimal local information contributes to the successful reconstruction of naturalistic images even in challenging conditions.

KEYWORDS

fast vision, image recognition, information-optimal local features, visual saliency, image occlusion

Introduction

In the real world, humans are constantly exposed to partially occluded objects, which the visual system must analyze and recognize very quickly for survival purposes. Thus, in real scenes, the visual system copes with the recognition of incomplete images, whose mechanisms are still not completely understood. Many studies have demonstrated that humans can successfully recognize fragmented images (Brown and Koch, 2000; Murray et al., 2001; Johnson and Olshausen, 2005; Ullman et al., 2016; Tang et al., 2018), but most of them focus on the rules to solve the occlusion and on how the system fills the missing information. Instead, here we are not interested in understanding the mechanisms through which the visual system binds the fragments into a whole image. We rather focus on the identification of the most relevant fragments to be analyzed and on the extraction of salient local features within these fragments. Hence, we focus on the low-level stages of this process.

To explain the mechanisms of information selection, several models of visual search employ the concept of *saliency map*, a two-dimensional map that encodes the saliency of the objects in the visual scene (Itti et al., 1998). The saliency of an object depends on its physical properties (e.g., luminance contrast, contours' orientation, etc.) and reflects the ability of that object to pop out in the visual scene. Each object in the scene competes for selection and only salient elements, those carrying the higher amount of information to the visual system, are chosen for further processing (Fecteau and Munoz, 2006). In this view, saliency operates very rapidly through bottom-up mechanisms: Salient features automatically attract our attention without any voluntary effort. From a neural perspective, it has been suggested that the primary visual cortex could provide a saliency map relying on specific processing of the local elements (Li, 2002). Visual saliency can be also influenced by contextual factors (Treisman and Gelade, 1980; Itti and Koch, 2001). Indeed, many studies related the selection of salient features to top-down mechanisms, by applying a spatially defined and feature-dependent weighting to the different feature maps (Wolfe et al., 1989).

The principles driving salience and the relative contribution of local (Li, 2002; Zhang et al., 2020) and global cues (Oliva and Schyns, 1997; Itti et al., 1998) are still under debate. Global and local information are related to spatial frequency: low spatial frequencies carry information about the global contrast distribution, whereas high spatial frequencies mainly provide fine information about local details (Blakemore and Campbell, 1969; Webster and de Valois, 1985; Boeschoten et al., 2005; Kauffmann et al., 2014). Nevertheless, several past studies have explored the mechanisms of fast vision at different scales and stimulus durations, finding that both coarse and fine spatial information are simultaneously used in fast image categorization (Oliva and Schyns, 1997; Schyns and Oliva, 1999).

In the present study, we hypothesize that the perception of incomplete images in fast vision partly starts from the extraction of some specific local high-frequency salient features contained in the visible image fragments. To identify salient features, we follow the principle that visual saliency may be based on the amount of local information (Shannon, 1948), as proposed by the constrained maximum-entropy model for early visual feature extraction (Del Viva et al., 2013). This model is founded on the need for a strong data reduction that must be operated by the visual system at an early stage, in order to optimize and speed up the reconstruction of visual images (Attneave, 1954; Barlow, 1961; Marr and Hildreth, 1980; Marr, 1982; Atick and Redlich, 1990; Atick, 1992; Olshausen and Field, 1996; Zhaoping, 2006). This is necessary given the huge amount of input data and the limited amount of neural resources (Attwell and Laughlin, 2001; Lennie, 2003; Echeverri, 2006; Del Viva and Punzi, 2014).

According to their model (Del Viva et al., 2013), in order to compress information and provide a saliency map of the visual scene, at an early stage the visual system selects only a very limited number of visual features for further processing. The features selected (*optimal* features) are those that produce in the output the largest amount of entropy allowed by the given computing limitations of this early stage filter (constrained maximum-entropy). The limitations considered by this model are the number of features transmitted and the output bandwidth (i.e., bandwidth and storage occupancy). Optimizing for entropy, together with the strict limitations on the computing resources, allows the system to completely determine the choice of the features from the statistical distribution of the input data. The authors proposed that only these features, which are *optimal* carriers of information, are *salient* in fast vision and used to represent visual images (*sketches*). All the other features that do not fulfill constrained maximum-entropy optimization criteria (*non-optimal* features) are considered not salient and are not transmitted to the following processing stages. Thus, unlike other models of early data compression based on redundancy reduction (Olshausen and Field, 1996), this approach leads to a huge loss of information. This is unavoidable given the limitations of the brain's capacity, imposed by intrinsic energetic costs of neuronal activity and ecological limits to the number of neurons. The result is a fast, albeit heuristic, analysis of salient features in the visual scene. The implementation of the model on a set of black and white naturalistic images (i.e., depicting landscapes, animals, plants; Olmos and Kingdom, 2004), imposing strict limitations on the number of features and output bandwidth, led to the extraction of a set of *optimal* features, that, according to the model, are the only visual elements used to build the image *sketch*.

The reduction of input images to only two levels is a corollary of the central idea of compression by pattern filtering proposed by the model (Del Viva et al., 2013): The number of possible patterns, assumed to be a limited resource, increases exponentially with the number of allowed levels (that is 2n*N where n is the number of bits and N the number of pixels)—and so does the amount of computing needed to calculate them. Therefore, using a large number of gray levels in the model would be not only unpractical but also would defeat its very purpose of saving computational resources. For the same reason, the authors chose to implement the model by defining as a feature a 3 × 3-pixel image partition. Such a small size, corresponding to about 6 × 6 min of arc, also allows to target early visual processing stages. These are very likely the anatomical substrate of the hypothesized filter because data compression must be done very early in the visual stream to be effective. Although early visual structures comprise multiple cell types, with different receptive field sizes (Nassi and Callaway, 2009), here, for simplicity, a single small scale is considered. However, this small scale is consistent with receptive field sizes found in human V1, which are about 15′ in the fovea (Smith et al., 2001) and become progressively larger with eccentricity and through the hierarchy of visual areas (Zeki, 1978). At any rate, features of this size have been demonstrated to be still visually discernible by normal human subjects (Del Viva et al., 2013).

Sketches, obtained by retaining only *optimal* features in the digitized images, were presented very briefly to human observers and allowed very accurate discrimination of the original unfiltered images (higher than 80%), comparable even to that of showing the images themselves. Nevertheless, the loss of information was conspicuous: information contained in the sketches could reach 10% of the originals, compressing data by a factor of 40 (Del Viva et al., 2013).

The spatial structure of extracted features resembles the bar- and edge-like receptive fields found in primary visual cortices (Hubel and Wiesel, 1965), suggesting that these specific visual receptive fields represent the optimal way to transmit information in fast vision. In contrast, the features discarded by the model as *non-optimal* carriers of information have a uniform luminance structure (features with high bandwidth occupancy) or a "noisy" alternation of black and white pixels (features with high memory occupancy) (Del Viva et al., 2013).

In a further study, to assess the contribution of *optimal* local features to image discrimination, they were replaced with *non-optimal* features along the objects' contours in the sketch. The disruption of optimal local cues in the sketches caused a decrease in image discriminability, despite preserving the global structure, suggesting that the fine structure of the image plays a crucial role in the discrimination (Del Viva et al., 2016).

Very recently, further studies showed that indeed these *optimal* features are considered salient even if they are presented in isolation without a global or semantic context (Castellotti et al., 2021), and they are able to automatically attract covert and overt attention (Castellotti et al., 2022).

Here we explore whether these specific local features still play an important role in more natural settings, where all existing features are kept (optimal and non-optimal), but the overall available information is drastically reduced. For this purpose, we created images where only a few fragments are shown, and the remaining parts are covered by a gray mask. In this way, we obtain visual stimuli with the same properties as the original images, in which the features are spatially and structurally unaltered, but the overall available information is reduced. To find the essential information needed to discriminate a visual scene, we pushed the visual system to its limits: the stimuli had very few visible parts and short durations. Specifically, participants had to covertly attend to a few briefly presented small fragments (or just one fragment) of binarized images (Del Viva et al., 2013) and then use them to discriminate the underlying image (target) from another (distractor).

Observers could solve this task by matching the position of black and white parts of the fragmented image and the target (global information), without the need to analyze the internal content of the fragments. If this were the case, we would expect the performance to depend on fragments contrast. On the other hand, performance could be related to the *optimal* information contained in the fragments, as predicted by the reference model. In this case, we would expect performance to depend on the number of local *optimal* features contained in the fragments. With multiple fragments covert attention could potentially be directed toward one of them; for this reason, we also measured discrimination by showing just a single fragment. This allowed us to correlate correct responses to the specific local information and contrast.

We then repeated the same discrimination task randomly inverting the contrast of the target and/or the distractor image. The purpose of this manipulation is to reduce the contribution of global information, given by the position of black/white large areas, and bring out the contribution of high-frequency components that could be masked by the prevalence of positional cues in original-contrast images.

Before testing our main experimental hypothesis in the Main experiment, we conducted two Preliminary experiments to test the limits for the discrimination of our fragmented digitized images, shown for a very short time. In these experiments, we probed the size and number of the fragments to be used in the Main experiment.

Materials and methods

Observers

Twenty young volunteers took part in this study. Ten observers (mean age = 25.3 ± 1.8 years) participated in Preliminary experiment 1, and five of them (mean age = 25.2 ± 1.8 years) also participated in Preliminary experiment 2. Ten other observers (mean age = 26.5 ± 2.9 years), all different from those of the preliminary experiments, participated in the Main experiment. All observers had normal or corrected to normal vision and no history of visual or neurological disorders. All participants gave written informed consent before the experiments. The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the local ethics committee ("Commissione per l'Etica della Ricerca," University of Florence, 7 July 2020, n. 111).

Apparatus and set-up

The apparatus and set-up were the same for the Preliminary and the Main experiments. All stimuli were programed on an ACER computer running Windows 10 with Matlab 2018b, using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007). The experiment was displayed on a gamma-corrected CRT Silicon Graphics monitor (1,152 × 864 pixels resolution, 38.5 × 29.5 cm, 120 Hz refresh rate), subtending 38.5 × 29.5 degree of visual angle at a 57 cm viewing distance. All experiments were carried out in a completely dark room. Participants' manual responses were provided on a standard Dell keyboard.

Procedure and stimuli

Preliminary experiment 1

The experimental procedure is represented in **Figure 1A**. Each trial started with the presentation of a white fixation point (300 ms) on gray background (14 cd/m^2) followed by the brief presentation (25 ms) of one stimulus in the center of the screen. Stimuli were composed of a certain number of image fragments of different sizes, resulting in a kind of "covered" image, revealing only small visible parts to the observer (see the paragraphs below for stimuli details). Immediately after, a mask appeared for 500 ms, followed by two black-white images sequentially presented for 350 ms each. One of the two images corresponded to the fragmented "covered" image (*target*), while the other (*distractor*) was randomly extracted from the set of images used (see the paragraphs below for image details). At each trial, the target was randomly presented in the first or the second interval. Images in the task were randomly displaced diagonally by 10 pixels, either to the top-left, top-right, bottom-left, or bottom-right, with respect to the position of the fragmented "covered" image. This spatial shift was purposedly introduced to avoid exact spatial matching between stimulus and target image. Observers were required to discriminate the target in a two-interval forced choice task (2IFC), by pressing a computer key.

Stimuli were prepared starting from 327 1-bit black and white renditions of naturalistic images, extracted from a public database (Olmos and Kingdom, 2004). Images' size was 918 × 672 pixels,

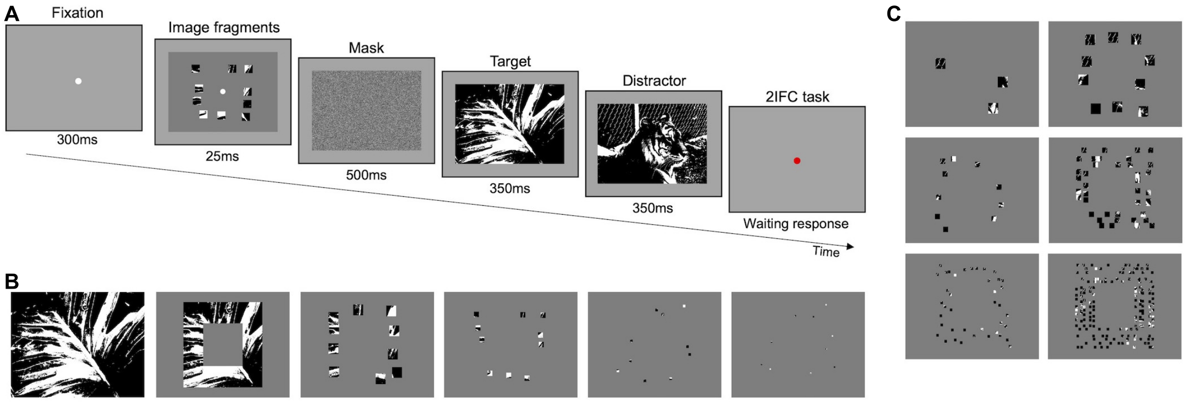


FIGURE 1
Preliminary experiments–Procedure and stimuli. **(A)** Representation of experimental paradigm. **(B)** Examples of stimuli used in Preliminary experiment 1. The first image is the control stimulus, the second is the "frame" stimulus, and the others show 10 fragments of decreasing size (in order: 7.5, 2, 0.47, and 0.12%), positioned within the frame. **(C)** Examples of stimuli used in Preliminary experiment 2. In the left column, fragments revealed 2% of the image area, and in the right column, fragments revealed 7.5% of the image area. Fragments' size in the images of each column decreases by fifty percent going from top to bottom; whereas fragments in the same row have the same size but vary in number.

subtending 32.4 × 23.7° of visual angle at 57 cm. The luminance of white, black, and medium gray was 35, 1, and 12 cd/m^2, respectively.

In Preliminary experiment 1, we measured discrimination as a function of the image's visible area. We used the following stimulus configurations: the whole image as a control (100% visible area, see **Figure 1B**—first panel); a squared "frame" comprised between 4.8° and 8.8° of eccentricity (35.8% visible area, see **Figure 1B**—second panel); ten image fragments revealing different fractions of image area: 7.5% (size of all fragments 2.4 × 2.4°), 2% (size of all fragments 1.2 × 1.2°), 0.47% (size of all fragments 0.6 × 0.6°) and 0.12% (size of all fragments 0.3 × 0.3°; see **Figure 1B**—third to sixth panels, respectively). In these cases, the rest of the image was covered by uniform gray pixels. For further examples of stimuli, see **Supplementary Figure 1**. For each area, image fragments were randomly selected from all possible combinations satisfying the following conditions: (i) They had to be comprised in the 4.8–8.8° eccentricity frame (stimuli presented within this eccentricity are well visible even if observers have to maintain fixation in the center, as shown with other tasks; see for example, Larson and Loschky, 2009; Staugaard et al., 2016); (ii) they had to be evenly distributed within the frame three fragments on the top and bottom sides of the frame, and two fragments on each lateral side; (iii) they could not overlap with each other. The chosen frame width guarantees that criteria (ii) and (iii) are met. For each image, five different fragments' configurations were created to minimize memory effects, for a total of 1,635 different stimuli for each area (see **Supplementary Figure 2**). A total of 3,000 trials per observer were run (300 trials for the control and frame conditions and 600 trials for each other condition). Each specific image configuration in each condition has been shown on average 1.2 times to each participant, preventing the association of a specific configuration of fragments to a target.

Preliminary experiment 2

Preliminary experiment 2 followed the same procedure as Preliminary experiment 1 (see **Figure 1A**). We measured discrimination as a function of the number of fragments of different sizes covering two different visible image areas (2 and 7.5%). The fragments were still positioned in the 4.8°−8.8° eccentricity frame. For 2% of the area we used: three 2.4 × 2.4° fragments (randomly distributed across the frame), ten 1.2 × 1.2° fragments (three fragments located on the top and bottom sides of the frame, and two fragments on the left and right sides), and 40 0.6 × 0.6° fragments (12 fragments located in the upper and lower side, and eight fragments in the left and right sides; see **Figure 1C**–left side panels, from top to bottom, respectively). For 7.5% of the area we used: 10 2.40 × 2.40° fragments (three fragments located on the top and bottom sides of the frame, and two fragments on the left and right sides), 40 1.2 × 1.2° fragments (12 fragments located on the top and bottom sides of the frame, and eight fragments on the left and right sides), and one 160 0.6 × 0.6° fragments (40 fragments located in the top, bottom, left, and right part of the image frame) (see **Figure 1C**–right side panels, from top to bottom, respectively). For further examples of stimuli, see **Supplementary Figure 3**. For each image, five different fragments' configurations were created, for a total of 1,635 different stimuli for each area (see **Supplementary Figure 2**). A total of 3,600 trials per observer were run (600 trials for each condition). Each specific image configuration in each condition has been shown on average 1.1 times to each participant.

Main experiment

The Main experiment follows the same procedure (2IFC) and used the same set of images (Olmos and Kingdom, 2004) as those of the Preliminary experiments 1 and 2, but participants were engaged in two different tasks: a task with original-contrast images and a task with randomly inverted-contrast images. In the first task, both the target and the distractor were digitized versions of the original images (as in **Figure 1A**). In the second task, in some randomly selected trials, the target and/or the distractor had their contrast inverted with respect to their original version (**Figure 2A**). Therefore, in some trials both the target and the distractor could be presented with their original or inverted contrast, while, in other trials, only one of them could have inverted contrast. With this manipulation, we aim at reducing the probability of solving the task by matching the position of black and white spots in the fragments to those in the images (see **Supplementary Figure 4**). Each image has been presented to each participant on average 37.7 times, either as a target or distractor.

In both tasks, the same conditions were tested. Stimuli consisted of one or ten fragments (see **Figure 2B**—first and second column,

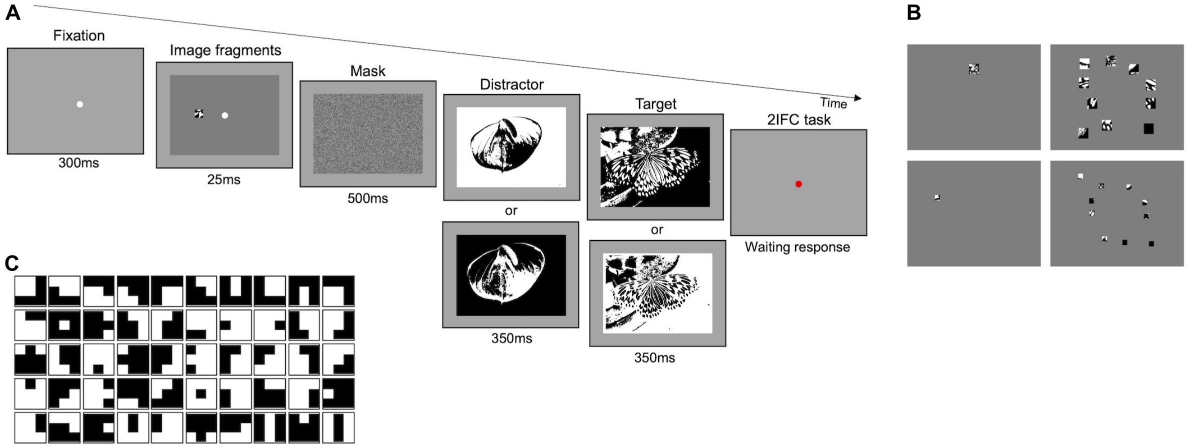


FIGURE 2
Main experiment–Procedure and stimuli. **(A)** Representation of experimental paradigm. The example illustrated in the upper row shows a trial where the distractor is presented with inverted contrast and the target with its original. The two panels below target and distractor indicate that in some trials either one or both of them can have a contrast inverted with respect to those shown above. **(B)** Examples of stimuli. Fragments in the images of each column are the same number but their size decreases by 50 percent from top to bottom; fragments in the images of each row have the same size but vary in number (1 or 10). **(C)** Set of *optimal* features. Set of 50 3 × 3-pixels features extracted by applying the constrained maximum entropy model to black and white images (Del Viva et al., 2013).

respectively) with different sizes: 2.4 × 2.4° and 1.2 × 1.2° (see **Figure 2B**–first and second row, respectively). The total area revealed by these fragments was 0.2 and 0.75% with one fragment, 2 and 7.5% with ten fragments. The characteristics of the stimuli (luminance, fragments distribution, and eccentricity) were the same as those used for Preliminary experiments 1 and 2. For further examples of stimuli, see **Supplementary Figure 5**. In the condition with 10 fragments, for each image, five different fragments' configurations were created, for a total of 1,635 different stimuli for each area (see **Supplementary Figure 2**). In the condition with 1 fragment four/five different configurations were created, for a total of 1,144 and 1,253 different stimuli for 0.2 and 0.75% area, respectively (see **Supplementary Figure 6**). In the Main experiment, each observer performed 2,400 trials in total: 1,200 trials in the task with original-contrast images (300 trials for each stimulus condition), and 1,200 trials in the task with randomly inverted-contrast images (300 trials for each stimulus condition). Each specific image configuration in each condition has been shown on average 1.1 times to each participant.

Data processing and statistical analysis

In all experiments, we measured the percentage of correct responses of each observer in each condition of visible area.

In Preliminary experiments 1 and 2, non-parametric one-way repeated-measures ANOVAs (Friedman's tests) with Conover *post hoc* comparisons (Bonferroni correction) were used to test differences between averaged performances across conditions. In Preliminary experiment 1, we also performed a one-sample Wilcoxon signed-rank test to assess whether the averaged performance in the condition with the smallest visible image area was still above the chance level (i.e., statistically different from 50%).

In the Main experiment, non-parametric two-way repeated-measures ANOVAs (Durbin tests) with Conover *post hoc* comparisons (Bonferroni correction) were used to test differences between average participants' performances in each condition of visible area in the original vs. inverted contrast tasks.

In addition, all observers' data were pooled together to calculate the performance as a function of fragments' contrast and signal-to-noise ratio (SNR) in each condition of visible area.

We calculated the Weber contrast of the fragment as follows: We first averaged the pixel values within the fragment (black = 0, white = 255), then this averaged value was subtracted from the background value (gray = 127), and finally the absolute value of the ratio between the result of the subtraction and the background was calculated. In the stimuli containing ten fragments, the average contrast of the fragments was considered. The performance was then analyzed as a function of Weber contrast (bins of 0.2 each).

To quantify the *saliency* of each fragment we calculated the signal-to-noise ratio (SNR), that is the number of *optimal* features, predicted salient by the reference model, over the total number of features. Specifically, we considered a set of 50 *optimal* features, 3 × 3 pixel large (see **Figure 2C**), each subtending ~0.1 × 0.1° of visual angle (about 12 c/deg spatial frequency). This specific set of *optimal* features has been proven to be salient for humans in previous works (Del Viva et al., 2013; Castellotti et al., 2022, 2021). In the stimuli containing ten fragments, the average SNR of the fragments was considered. The performance was then analyzed as a function of SNR (bins of 0.05 each).

For each SNR bin, we calculated the average contrast of fragments with the standard error. The Pearson linear-correlation coefficient between SNR and contrast was then calculated.

Given the strong correlation between fragments' contrast and SNR, to quantify their relative contribution to the performance, we created a new variable by subtracting, in each trial, the standardized values from each other (SNR—contrast).

Data from all conditions of visible area (7.5, 2, 0.75, and 0.2%) were pooled together and GLMMs with a binomial error structure were performed. In the task with original contrast images, the model included three fixed factors: (i) SNR-contrast difference (standardized); (ii) target order presentation, to test whether the performance depended on the fact that the target was in the first vs. second interval; (iii) image repetition number (i.e., the frequency of occurrence of each image as target or distractor), to control for possible effects of visual memory. Participants and stimuli were included as random effects. In the task with randomly inverted-contrast images an additional fixed factor was included: (iiii) target contrast inversion, to test whether the performance changed in the trials where the target was presented with original or inverted contrast.

We then compared (*z*-tests) the probability of correct responses (with binomial standard deviations) between the task with original-contrast images and the one with random contrast inversion. This was done separately for the trials where the target had original contrast and for those where the target had inverted contrast.

Finally, a GLMM was run in the task with randomly contrast-inverted images including only the trials where the target had original contrast.

Results

Preliminary experiment 1

Average performance in Preliminary experiment 1 ($n = 10$) is reported in **Figure 3A**. As expected, the percentage of correct responses increases with the size of the image fragments (i.e., the amount of visible area of the image). On average, observers' performance ranges from 55% for the smallest visible area to 83% when the full image is shown (100% area). Particularly, observers gave 54.5 ± 1.03% (SE) correct responses at 0.12% of visible area, 58.3 ± 1.7% at 0.47%, 62 ± 1.6% at 2%, 65.8 ± 1.9% at 7.5%, 75.4 ± 2.4% at 35.8, 83.1 ± 2.5% at 100%. Friedman's test showed a main effect of the visible area [$\chi^2(5) = 45.3$, $p < 0.001$, $W = 0.46$]. All Conover *post hoc* comparisons (Bonferroni correction) are reported in **Supplementary Table 1**.

The average performance obtained by showing the smallest image area also resulted statistically different from 50% [$Z\ (9) = 55$, $p = 0.002$], showing that observers are able to discriminate an image based on very little information.

Preliminary experiment 2

In Preliminary experiment 2 ($n = 5$), we compared the observers' performance when the same amount of image area is revealed by showing a different number of fragments of different sizes. Performances are reported in **Figure 3B**. For both areas tested (2 and 7.5%), the percentage of correct responses tends to be greater with few big fragments than with more small fragments, even if none of the results are statistically significant. When the size of the patches remains constant but their number increases, thus revealing a bigger amount of image area to the observers, the performance slightly increases in all conditions, although not significantly. Specifically, when the percentage of the revealed image area is 2%, average performance is 66.3 ± 2.1% (SE) with three 2.40 × 2.40° fragments, 63.03 ± 2.3% with ten 1.20 × 1.20° fragments, and 61.1 ± 1.5% with forty 0.6 × 0.6° fragments. When the percentage of the revealed image area is 7.5%, average performance is 67.6 ± 2.8% with ten 2.40 × 2.40° fragments, 65.9 ± 2.2% with forty 1.20 × 1.20° fragments, and 62.1 ± 2.8% with one hundred and sixty 0.6 × 0.6° fragments.

Main experiment

In the Main experiment ($n = 10$), we first analyzed the percentage of correct discrimination in the two tasks. In the task with original-contrast images (**Figure 4A**), when ten fragments are presented, observers' discrimination is 63.3 ± 1.8% (SE) for 2% area and 68.8 ± 2.5% for 7.5% area (**Figure 4A**—left panel). With one single fragment, the average observers' performance is 60.7 ± 2% at 0.2% area and 64.3 ± 1.6% at 0.75% area (**Figure 4A**—right panel). In the task with randomly inverted-contrast images (**Figure 4B**), with ten fragments discrimination performance is 61.1 ± 1.8% at 2% area and 66.7 ± 2.2% at 7.5% area (see **Figure 4B**—left panel). With one single fragment, the average observers' performance is 58.3 ± 1.3% at 0.2% area and 63.6 ± 2.1% at 0.75% area (**Figure 4B**—right panel). Durbin test between performances with original- vs. randomly inverted-contrast images confirmed the effect of visible area [$\chi^2(1) = 9.2$, $p = 0.002$, $W = -20$] but no statistical differences emerged across the two tasks [$\chi^2(1) = 0.2$, $p = 0.61$]. This suggests that, even if in some trials of this task there is no correspondence between the contrast of the fragments and that of the target image, the overall performance is comparable to that obtained in the task with original-contrast images.

We then investigated to what extent the performance depended on the *saliency* of the local high-frequency features contained in the fragments presented (as predicted by the constrained maximum-entropy model), or on the global luminance information (Weber *contrast*). Firstly, we calculated performance as a function of SNR and contrast separately. In the task with original contrast images, performance does not depend on SNR, and it does not seem to be related to fragments' contrast as well, although there is a tendency to increase with contrast with multiple fragments (**Supplementary Figures 7A–B**). Instead, in the task with randomly inverted-contrast images, the performance is higher for lower contrasts and decreases for higher contrasts, whereas it increases from lower to higher SNR (**Supplementary Figures 7C–D**).

Note however that fragments' contrast and SNR are negatively correlated (**Figure 5**; 7.5% area: $r = -0.63$, $p < 0.001$; 2% area: $r = -0.72$, $p < 0.001$; 0.75% area: $r = -0.60$, $p < 0.001$; 0.2% area: $r = -0.69$, $p < 0.001$). This correlation depends on the nature of the fragments and the way the two variables have been calculated: fragments with lower contrast are those containing a higher number of optimal features (high SNR), because high SNR reflects into a textured stimulus, and averaging alternations of many black and white pixels, leads to low Weber contrast. On the other end, fragments with higher contrast are those with large black/white parts and therefore contain a few optimal features (see **Supplementary Figure 8**). Note that the maximum SNR in the case of ten fragments (0.2) is lower than for one fragment (0.3) because, being the contrast mediated across ten different parts, the probability of having fragments with large black and white parts (and consequently low SNR) is higher.

Since the correlations between SNR and contrast are quite high, in the following analysis we used the difference between standardized SNR and contrast, instead of considering them as two separate variables. In this way, the contributions of SNR and contrast to the performance can be separated. Moreover, in

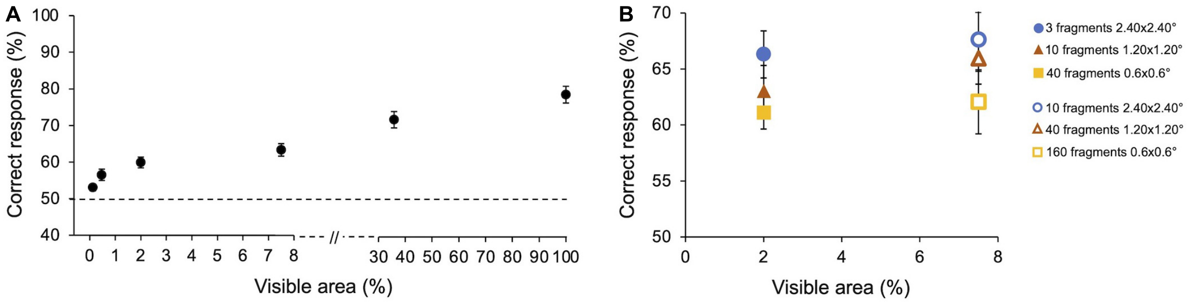


FIGURE 3
Preliminary experiments–Results. **(A)** Performance as a function of images' visible area. Performance averaged across participants (*n* = 10) with SE. Observers performed 3,000 trials in total (300 trials for 100% area and frame conditions and 600 trials for each other areas condition). **(B)** Performance as a function of the number and size of image fragments. Performance averaged across participants (*n* = 5) with SE. Two image areas have been tested: Filled symbols indicate fragments revealing 2% of the area; empty symbols indicate fragments revealing 7.5% of the area. Symbols with the same shape indicate a different number of fragments of the same size. Observers performed 3,600 trials in total (600 trials for each condition).

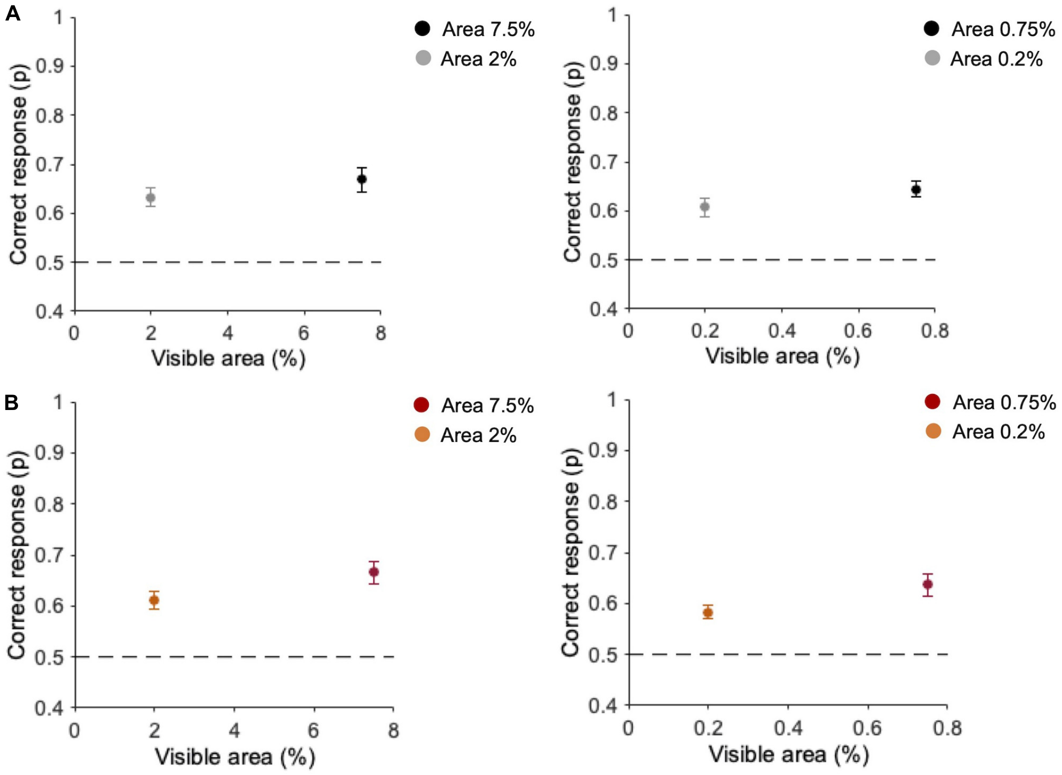


FIGURE 4
Performance for different areas and number of fragments. **(A)** Task with original-contrast images. **(B)** Task with randomly inverted-contrast images. Left panels: average performance (*n* = 10) for ten fragments (2 and 7.5% of area); Right panels: average performance (*n* = 10) for one fragment (0.2 and 0.75% of area). Errors are SE across participants. Observers performed 2,400 trials in total (300 trials for each condition).

a 2IFC task, the order of target presentation might affect the performance, as well as the frequency of occurrence of each image: repeated presentations of the same image as target or distractor might induce visual learning of the images. For the task with original contrast images, we then performed a GLMM with three fixed factors: SNR-contrast difference (standardized), target order presentation, and image repetition number. Participants and stimuli were included as random effects. The GLMM reveals no effect of the difference between standardized SNR and contrasts [$\chi^2(1) = 0.24$, $p = 0.62$], but a main effect of order [$\chi^2(1) = 9.1$, $p = 0.002$] and image repetition number [$\chi^2(1) = 19.2$, $p < 0.001$] emerges. Contrasts and marginals means are reported in **Supplementary Table 2**.

Overall, these results indicate that, in the task with original-contrast images, the performance does not depend on SNR (as shown in **Figure 6A**), and it does not seem to be related to fragments' contrast either (although there is a tendency to increase with contrast with multiple fragments; see **Supplementary Figures 7A–B**). Given our hypotheses, we argue that in this condition observers do not rely on local cues and possibly use the position of black and white spots to solve the task. This hypothesis seems to be further supported by the fact that the performance is higher when the target is presented in the

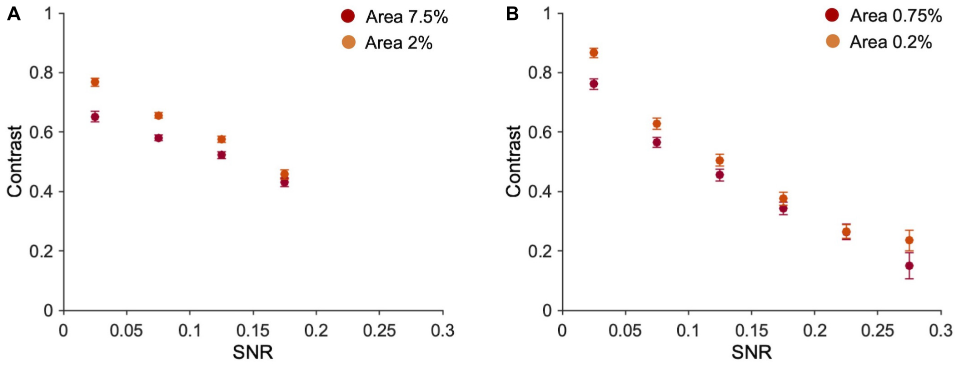


FIGURE 5
Fragments' contrast vs. their SNR. **(A)** Average contrast of ten fragments vs. their average SNR (bins of 0.05 each). Number of occurrences in each bin (from the first to the last bin), 7.5% area: 345, 1,380,783,491; 2% area: 497,1202,857,444. **(B)** Average contrast vs. averaged SNR of one fragment (bins of 0.05 each). Error bars are standard errors. Number of occurrences in each bin (from the first to the last bin) = 0.75% area: 834,873,586,425,235,47; 2% area: 748,764,620,440, 308,120.

first interval of the 2IFC. Indeed, the match between the fragments and the corresponding image is easier if the target is temporally closer and its presentation is not interspersed with the appearance of the distractor.

We then performed the same analysis in the task with randomly inverted-contrast images (**Figure 6B**), used to reduce the contribution of positional global cues and to bring out the contribution of high-frequency optimal features (see **Supplementary Figure 4**). In this task, an additional factor was included in the GLMM. Considering all visible area conditions (12,000 trials in total), due to the random nature of inversion, the target contrast alone was inverted in 24.5% of trials, the distractor contrast alone was inverted in 25.2% of trials, the contrasts of both the target and distractor were inverted in 22.8% of trials, and the contrasts of both target and distractor were kept original in 27.4% of trials. In principle, these different target conditions could affect performance. The GLMM analysis was thus performed with four fixed factors (standardized SNR-contrast difference, target order presentation, image repetition number, and target contrast inversion) and two random effects: participants and stimuli. Contrasts and marginal means are reported in **Supplementary Table 3**. The analysis shows a significant effect of SNR-contrast difference [$\chi^2(1) = 128.4, p < 0.001$] on performance. Indeed, performance increases with this difference (**Figure 6B**), suggesting that SNR prevails over contrast in driving the performance. The target order factor is instead not statistically significant [$\chi^2(1) = 0.07$, $p = 0.78$], meaning that the performance does not change whether the target image is shown in the first or the second interval of the 2IFC task. These results confirm further our hypothesis that, in this condition, participants change their strategy: They do not rely on positional cues anymore, but rather they use local information, therefore target order does not affect the performance. Again, the analysis reveals an effect of the image repetition number [$\chi^2(1) = 36.2$, $p < 0.001$]. The target contrast inversion factor is also statistically significant [$\chi^2(3) = 45.5$, $p < 0.001$]. Indeed, the performance with original-contrast target (65 ± 0.006%) is higher than with inverted-contrast target (60 ± 0.006%).

Interestingly, the performance in the task with randomly inverted-contrast images in the trials with original-contrast target is also higher than that obtained in the task with original-contrast images (63 ± 0.004%; $z = 2$, $p = 0.04$), although these two conditions are exactly the same.

The GLMM analysis, including only the trials with original-contrast target of the task with randomly inverted-contrast images, reveals a main effect of the difference between SNR–contrast [$\chi^2(1) = 33.9$, $p < 0.001$; see **Supplementary Figure 9**], and of image repetition number [$\chi^2(1) = 18.4$, $p < 0.001$], but there is no effect of target presentation order [$\chi^2(1) = 0.31$, $p = 0.58$]. Contrasts and marginals means are reported in **Supplementary Table 4**. These results are compatible with those found when considering all trials, independently of target contrast inversion (see **Figure 6B**). On the other end, these results are different from those found in the task with original-contrast images (see **Figure 6A**), although these two conditions are exactly the same. See the Discussion section for the interpretation of these results.

Discussion

In the present work, we investigated the visual system's ability to quickly discriminate a scene, based on the salience of high-frequency local visual features.

Over the years, different studies have argued that the selection of relevant local elements is based on the simultaneous processing of different visual properties at multiple spatial scales, then combined into a single saliency-map (Itti et al., 1998; Itti and Koch, 2001; Torralba, 2003). However, these models do not consider the amount of computing power required by each parallel process. Our reference model, instead, takes into account the system's computational costs. Considering the finest spatial scale as the most computationally demanding part of the processing and the need for fast analysis, the model applies a lossy data compression algorithm to images at a fine spatial scale (Del Viva et al., 2013). The result of this process is the extraction of a limited number of informative high-frequency visual features, that are used for fast image discrimination and to drive bottom-up attention (Castellotti et al., 2022, 2021).

Before investigating their role in fast discrimination of fragmented images, often presented to the visual system due to occlusions, we showed that observers can discriminate an image presented only for 25 ms even when it's almost totally occluded. As expected, correct discrimination increases with the visible area, but is still possible with very little information (0.12%). These findings

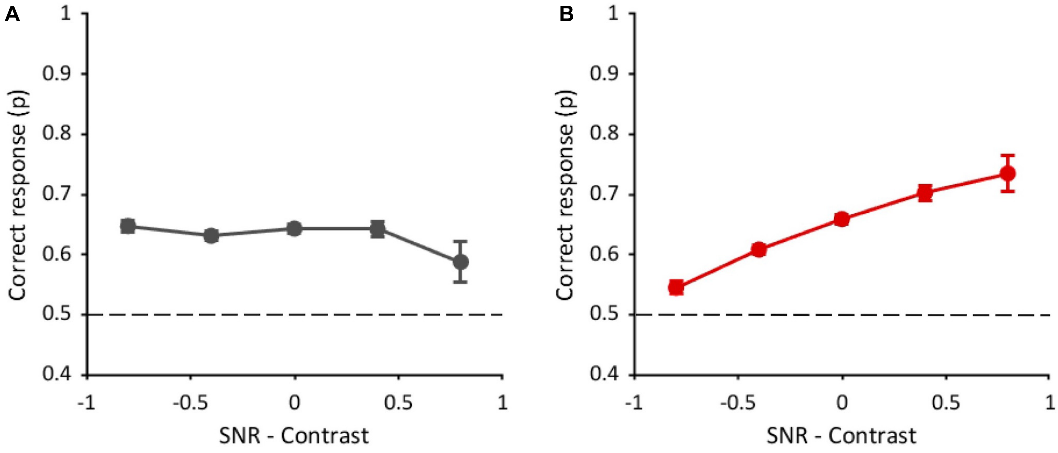


FIGURE 6
Performance as a function of the difference between standardized SNR and contrast. **(A)** Task with original-contrast images. **(B)** Task with randomly inverted-contrast images. Data from all observers (n = 10) and all conditions of visible area (7.5, 2, 0.75, and 0.2%) are pooled together. Errors are binomial standard deviation. Dashed lines represent chance level.

confirm that humans are very skilled in fast visual discrimination, as already broadly demonstrated (for a review, see Serre et al., 2007). Note however that we pushed the visual system's capacity to its limit, by showing images for the minimum duration necessary for a visual stimulus to reach primary cortical visual structures (Grill-Spector et al., 2000; Kirchner and Thorpe, 2006) and by using a paradigm that is known to be challenging for the observers (i.e., 2IFC tasks lead to higher error than 2AFC, Jäkel and Wichmann, 2006); This might explain why observers did not reach top performance even when the full image is displayed (100% area). Despite this, the minimal percentage of visible area needed to perform the task is much lower (0.12%) than that found in previous studies. For example, Tang et al. (2018) conducted an experiment similar to ours, with occluded or partially visible images presented for different durations, finding that in 25 ms observers robustly recognized objects when they were rendered <15% visible (Tang et al., 2018). The higher performance with a smaller visible area found here could be explained by the different tasks involved: their participants had to choose the right association between the occluded content and five different label options, while ours discriminate between two images.

We also investigated which factors mostly influence the correct discrimination of occluded pictures. That is, we studied whether, with the same amount of visible area, discrimination depends more on the number of visible fragments or on their size. Results show a slight (not significant) preference for a few large fragments, rather than for many small parts. This is somewhat unexpected. However, some have hypothesized that perceptual systems suffer from overload, so the higher the perceptual load of current information, the lower the ability to perceive additional information (Greene et al., 2017). Here a low number of fragments could produce a lower cognitive load (Xing, 2007; Nejati, 2021), hence better performance.

In the Main experiment, we investigated the role of the high-frequency model-predicted *optimal* features in fragmented image discrimination by quantifying the saliency of the fragments as the ratio of *optimal* features over the total number of features they contain. That is, the question is whether observers focus on the local internal content of the fragments and use embedded optimal features to discriminate the target, or whether they covertly attend to the global contrast information (low frequency). Indeed, since we use black and white stimuli and a 2IFC discrimination task, observers could simply solve the task by matching the position of black and white parts of the fragmented image and the target, without the need to analyze the internal content of the patches.

When low frequencies can be used to perform the task (original contrast), the performance does not depend on the number of optimal features contained in the fragments, rather there is a slight tendency to increase with fragments contrast (particularly when ten fragments are shown). These results suggest that in this condition observers do not use local information but possibly use the fragments' global luminance distribution. This hypothesis is further supported by the evidence that, only in the task with original-contrast images, the performance increases if the target is shown in the first interval of the 2IFC task. Indeed, we can assume that the match between the position of the black and white parts of the fragmented image and the target is easier if the latter is temporally closer to the stimulus and there is no other image before it.

A higher performance in the task with original-contrast images than in the task with random contrast inversion would be expected, since, in the former, positional cues can always be used. The fact that the performances in the two tasks are similar suggests that, when the contribution of global information is decreased (random inversion of contrast), observers rely on a different kind of information to discriminate the scene. In fact, we found that the probability of correct discrimination increases with the number of optimal features in the fragments, both with one and ten fragments, indicating that observers' responses in the task with random inversion of contrast are based on the local content of the fragments. This change of strategy is further supported by the evidence that, in this condition, the performance does not depend on the target order of presentation. We argue that, since observers do not base their choice on positional cues, it doesn't matter anymore if the target is presented in the first or in the second interval.

In the task with randomly inverted-contrast images in some trials the target still has the original contrast, therefore the global luminance structure of the fragments could still drive discrimination. Interestingly, considering only these specific trials, the performance is even higher than that obtained in the task where only original-contrast images are used, even though the two conditions are exactly the same. More importantly, correct responses depend on the number of optimal features in the fragments, and they are independent of target order, unlike in the task with original-contrast images. These results confirm that the contrast manipulation we applied in this task can change the observers' strategy. In this condition, participants seem to use both global and local information reaching a higher performance than when they rely only on global information. We, therefore, conclude that when less global information is available, local information plays a crucial role.

Note that the set of optimal features comprises spatial structures with both contrast polarities; this could explain why the inversion of contrast does not affect discrimination based on local information. The insensitivity to contrast inversion (Baylis and Driver, 2001; Niell and Stryker, 2008) found in V1 complex cells, together with the similarity of spatial structure between model-predicted optimal features and the bar and edge-like V1 receptive fields (Hubel and Wiesel, 1965), strongly suggests that these cells represent the optimal way to transmit information in fast vision. This also highlights the strong predictive power of the constrained maximum-entropy model.

Overall, our findings suggest that local and global analyses interact in fast image processing and that the contribution of the high-frequency optimal features significantly emerges when the visual system is tested in very challenging conditions. This means that local information, when derived from maximum-entropy optimization criteria coupled with strict computational limitations, allows fast image discrimination even when the information about the scene is drastically reduced.

This fast local extraction of salient features must be operated very early in the visual pathway (Li, 2002; Del Viva et al., 2013), and integrated into a global percept at later visual stages. Indeed, in real scenes the visual system "goes beyond the information given" in a local region (Meng and Potter, 2008) and fills in the missing information of occluded images by binding the visible image fragments (Bruno et al., 1997; Johnson and Olshausen, 2005; Meng and Potter, 2008). Also, in daily life, the *a priori* knowledge of the objects helps the visual system in image recognition (Pinto et al., 2015; Stein and Peelen, 2015). Long-term memory, which is capable of storing a massive number of details from the images (Brady et al., 2008), contributes as well. Visual learning effects also occurred in our experiment, since the performance is affected by repeated presentations of the same image. This indicates that participants might have become acquainted with image details, revealing that there are some memory effects at play. Studies of the mechanisms of recognition of incomplete images have also developed information-statistical approaches, the concepts of the extraction of the signal from noise, and models of matched filtration (for a review, see Shelepin et al., 2009).

To conclude, our study confirm that local visual saliency can be determined by the amount of information that local features carry about the visual scene weighed with their processing costs for the system, as predicted by the reference model (Del Viva et al., 2013). What cannot be ignored is the fact that while viewing a scene, humans make eye movements several times per second. Considering these results as a starting point for further studies, it would be interesting to investigate whether saccades are directed toward the most informative areas, represented by the optimal features predicted by our reference model, to reconstruct the image.

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://doi.org/10.5281/zenodo.7096390.

Ethics statement

The studies involving human participants were reviewed and approved by the Local Ethics Committee: "Commissione per l'Etica della Ricerca," University of Florence, 7 July 2020, no. 111. The patients/participants provided their written informed consent to participate in this study.

Author contributions

SC and OD participated in the experiment programing, data collection, statistical analyses, and manuscript writing and review. MMDV participated in the project ideation, statistical analyses, and manuscript writing and review. All authors contributed to the article and approved the submitted version.

Funding

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program (grant agreement no 832813 GenPercept "Spatio-temporal mechanisms of generative perception").

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2023.1049615/full#supplementary-material

References

Atick, J. J. (1992). Could information theory provide an ecological theory of sensory processing? *Netw. Comput. Neural Syst.* 22, 4–44. doi: 10.1088/0954-898X_3_2_009

Atick, J. J., and Redlich, A. N. (1990). Towards a theory of early visual processing. *Neural Comput.* 2, 308–320. doi: 10.1162/neco.1990.2.3.308

Attneave, F. (1954). Some informational aspects of visual perception. *Psychol. Rev.* 61, 183–193. doi: 10.1037/h0054663

Attwell, D., and Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metab.* 21, 1133–1145. doi: 10.1097/00004647-200110000-00001

Barlow, H. (1961). Possible principles underlying the transformations of sensory messages. *Sens. Commun.* 1, 217–234. doi: 10.7551/mitpress/9780262518420.003.0013

Baylis, G. C., and Driver, J. (2001). Shape-coding in IT cells generalizes over contrast and mirror reversal, but not figure-ground reversal. *Nat. Neurosci.* 4, 937–942. doi: 10.1038/nn0901-937

Blakemore, B. C., and Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *J. Physiol.* 203, 237–260. doi: 10.1113/jphysiol.1969.sp008862

Boeschoten, M. A., Kemner, C., Kenemans, J. L., and Van Engeland, H. (2005). The relationship between local and global processing and the processing of high and low spatial frequencies studied by event-related potentials and source modeling. *Cogn. Brain Res.* 24, 228–236. doi: 10.1016/j.cogbrainres.2005.01.021

Brady, T. F., Konkle, T., Alvarez, G. A., and Oliva, A. (2008). Visual long-termmemory has a massive capacity for object details. *Proc. Natl. Acad. Sci. U.S.A.* 105, 14325–14329. doi: 10.1073/pnas.0803390105

Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357

Brown, J. M., and Koch, C. (2000). Influences of occlusion, color, and luminance on the perception of fragmented pictures. *Percept. Mot. Skills* 90, 1033–1044. doi: 10.2466/pms.2000.90.3.1033

Bruno, N., Bertamini, M., and Domini, F. (1997). Amodal completion of partly occluded surfaces: Is there a mosaic stage? *J. Exp. Psychol. Hum. Percept. Perform.* 23, 1412–1426. doi: 10.1037/0096-1523.23.5.1412

Castellotti, S., Montagnini, A., and Del Viva, M. M. (2021). Early visual saliency based on isolated optimal features. *Front. Neurosci.* 15:645743. doi: 10.3389/fnins.2021.645743

Castellotti, S., Montagnini, A., and Del Viva, M. M. (2022). Information-optimal local features automatically attract covert and overt attention. *Sci. Rep.* 12:9994. doi: 10.1038/s41598-022-14262-2

Del Viva, M. M., and Punzi, G. (2014). The brain as a trigger system. *Proc. Sci.* 213:252. doi: 10.22323/1.213.0252

Del Viva, M. M., Punzi, G., and Benedetti, D. (2013). Information and perception of meaningful patterns. *PLoS One* 8:e69154. doi: 10.1371/journal.pone.0069154

Del Viva, M. M., Punzi, G., and Shevell, S. K. (2016). Chromatic information and feature detection in fast visual analysis. *PLoS One* 11:e0159898. doi: 10.1371/journal.pone.0159898

Echeverri, E. (2006). Limits of capacity for the exchange of information in the human nervous system. *IEEE Trans. Inf. Technol. Biomed.* 10, 803–808. doi: 10.1109/TITB.2006.879585

Fecteau, J. H., and Munoz, D. P. (2006). Salience, relevance, and firing: A priority map for target selection. *Trends Cogn. Sci.* 10, 382–390. doi: 10.1016/j.tics.2006.06.011

Greene, C. M., Murphy, G., and Januszewski, J. (2017). Under high perceptual load, observers look but do not see. *Appl. Cogn. Psychol.* 31, 431–437. doi: 10.1002/acp.3335

Grill-Spector, K., Kushnir, T., Hendler, T., and Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nat. Neurosci.* 3, 837–843. doi: 10.1038/77754

Hubel, D. H., and Wiesel, T. N. (1965). Receptive archi- tecture in two nonstriate visual areas (18 and 19) of the cati. *J. Neurophysiol.* 28, 229–289.

Itti, L., and Koch, C. (2001). Computational modelling of visual attention. *Nat. Rev. Neurosci.* 2, 194–203. doi: 10.1038/35058500

Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 20, 1254–1259. doi: 10.1109/34.730558

Jäkel, F., and Wichmann, F. A. (2006). Spatial four-alternative forced-choice method is the preferred psychophysical method for naïve observers. *J. Vis.* 6, 1307–1322. doi: 10.1167/6.11.13

Johnson, J. S., and Olshausen, B. A. (2005). The recognition of partially visible natural objects in the presence and absence of their occluders. *Vision Res.* 45, 3262–3276. doi: 10.1016/j.visres.2005.06.007

Kauffmann, L., Ramanoël, S., and Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Front. Integr. Neurosci.* 8:37. doi: 10.3389/fnint.2014.00037

Kirchner, H., and Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Res.* 46, 1762–1776. doi: 10.1016/j.visres.2005.10.002

Kleiner, M., Brainard, D. H., Pelli, D. G., Broussard, C., Wolf, T., and Niehorster, D. (2007). What's new in psychtoolbox-3? *Perception* 36, 1–16.

Larson, A. M., and Loschky, L. C. (2009). The contributions of central versus peripheral vision to scene gist recognition. *J. Vis.* 9:6.1–16. doi: 10.1167/9.10.6

Lennie, P. (2003). The cost of cortical computation. *Curr. Biol.* 13, 493–497. doi: 10.1016/S0960-9822(03)00135-0

Li, Z. (2002). A saliency map in primary visual cortex. *Trends Cogn. Sci.* 6, 9–16. doi: 10.1016/S1364-6613(00)01817-9

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* San Francisco, CA: W. H. Freeman. doi: 10.1016/0022-2496(83)90030-5

Marr, D., and Hildreth, E. (1980). Theory of edge detection. *Proc. R. Soc. Lond. B Biol. Sci.* 207, 187–217. doi: 10.1098/rspb.1980.0020

Meng, M., and Potter, M. C. (2008). Detecting and remembering pictures with and without visual noise. *J. Vis.* 8, 1–10. doi: 10.1167/8.9.7

Murray, R. F., Sekuler, A. B., and Bennett, P. J. (2001). Time course of amodal completion revealed by a shape discrimination task. *Psychon. Bull. Rev.* 8, 713–720. doi: 10.3758/BF03196208

Nassi, J. J., and Callaway, E. M. (2009). Parallel processing strategies of the primate visual system. *Nat. Rev. Neurosci.* 10, 360–372. doi: 10.1038/nrn2619

Nejati, V. (2021). Effect of stimulus dimension on perception and cognition. *Acta Psychol.* 212:103208. doi: 10.1016/j.actpsy.2020.103208

Niell, C. M., and Stryker, M. P. (2008). Highly selective receptive fields in mouse visual cortex. *J. Neurosci.* 28, 7520–7536. doi: 10.1523/JNEUROSCI.0623-08.2008

Oliva, A., and Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuliings of their distinctive properties. in the object recognition literature, scene. *Cogn. Psychol.* 34, 72–107.

Olmos, A., and Kingdom, F. A. A. (2004). A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception* 33, 1463–1473. doi: 10.1068/p5321

Olshausen, B. A., and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381, 607–609. doi: 10.1038/381607a0

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

Pinto, Y., van Gaal, S., de Lange, F. P., Lamme, V. A. F., and Seth, A. K. (2015). Expectations accelerate entry of visual stimuli into awareness. *J. Vis.* 15, 1–15. doi: 10.1167/15.8.13

Schyns, P. G., and Oliva, A. (1999). Dr. Angry and Mr. Smile: When categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition* 69, 243–265. doi: 10.1016/S0010-0277(98)00069-9

Serre, T., Kreiman, G., Kouh, M., Cadieu, C., Knoblich, U., and Poggio, T. (2007). A quantitative theory of immediate visual recognition. *Prog. Brain Res.* 165, 33–56. doi: 10.1016/S0079-6123(06)65004-8

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 623–656. doi: 10.1002/j.1538-7305.1948.tb00917.x

Shelepin, Y. E., Chikhman, V. N., and Foreman, N. (2009). Analysis of the studies of the perception of fragmented images: Global description and perception using local features. *Neurosci. Behav. Physiol.* 39, 569–580. doi: 10.1007/s11055-009-9171-1

Smith, A. T., Singh, K. D., Williams, A. L., and Greenlee, M. W. (2001). Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cereb. Cortex* 11, 1182–1190. doi: 10.1093/cercor/11.12.1182

Staugaard, C. F., Petersen, A., and Vangkilde, S. (2016). Eccentricity effects in vision and attention. *Neuropsychologia* 92, 69–78. doi: 10.1016/j.neuropsychologia.2016.06.020

Stein, T., and Peelen, M. V. (2015). Content-specific expectations enhance stimulus detectability by increasing perceptual sensitivity. *J. Exp. Psychol. Gen.* 144, 1089–1104. doi: 10.1037/xge0000109

Tang, H., Schrimpf, M., Lotter, W., Moerman, C., Paredes, A., Caro, J. O., et al. (2018). Recurrent computations for visual pattern completion. *Proc. Natl. Acad. Sci. U.S.A.* 115, 8835–8840. doi: 10.1073/pnas.1719397115

Torralba, A. (2003). Contextual priming for object detection. *Int. J. Comput. Vis.* 53, 169–191.

Treisman, A. M., and Gelade, G. (1980). A feature-integration theory of attention. *Cogn. Psychol.* 12, 97–136. doi: 10.1016/0010-0285(80)90005-5

Ullman, S., Assif, L., Fetaya, E., and Harari, D. (2016). Atoms of recognition in human and computer vision. *Proc. Natl. Acad. Sci. U.S.A.* 113, 2744–2749. doi: 10.1073/pnas.1513198113

Webster, M. A., and de Valois, R. L. (1985). Relationship between spatial-frequency and orientation tuning of striate-cortex cells. *J. Opt. Soc. Am. A* 2, 1124–1132. doi: 10.1364/JOSAA.2.001124

Wolfe, J. M., Cave, K. R., and Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *J. Exp. Psychol. Hum. Percept. Perform.* 15, 419–433. doi: 10.1037//0096-1523.15.3.419

Xing, J. (2007). "LNCS 4553–information complexity in air traffic control displays," in *Human-computer interaction. HCI applications and services. HCI 2007. Lecture notes in computer science*, Vol. 4553, ed. J. A. Jacko (Berlin: Springer), 797–806.

Zeki, S. M. (1978). Uniformity and diversity of structure and function in rhesus monkey prestriate visual cortex. *J. Physiol.* 277, 273–290. doi: 10.1113/jphysiol.1978.sp012272

Zhang, X., Sun, Y., Liu, W., Zhang, Z., and Wu, B. (2020). Twin mechanisms: Rapid scene recognition involves both feedforward and feedback processing. *Acta Psychol.* 208:103101. doi: 10.1016/j.actpsy.2020.103101

Zhaoping, L. (2006). Theoretical understanding of the early visual processes by data compression and data selection. *Network* 17, 301–334. doi: 10.1080/09548980600931995

frontiers | Frontiers in Human Neuroscience

TYPE Brief Research Report
PUBLISHED 21 February 2023
DOI 10.3389/fnhum.2022.1052604

OPEN ACCESS

EDITED BY
Elena Nava,
University of Milano-Bicocca, Italy

REVIEWED BY
Nico Bast,
Goethe University Frankfurt, Germany
Makoto Wada,
National Rehabilitation Center
for Persons With Disabilities, Japan

*CORRESPONDENCE
Jennifer B. Wagner
✉ jen.wagner@alum.mit.edu

SPECIALTY SECTION
This article was submitted to
Sensory Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 24 September 2022
ACCEPTED 16 December 2022
PUBLISHED 21 February 2023

CITATION
Soker-Elimaliah S, Lehrfield A,
Scarano SR and Wagner JB (2023)
Associations between the pupil light
reflex and the broader autism
phenotype in children and adults.
Front. Hum. Neurosci. 16:1052604.
doi: 10.3389/fnhum.2022.1052604

Associations between the pupil light reflex and the broader autism phenotype in children and adults

Sapir Soker-Elimaliah[1,2,3], Aviva Lehrfield[4],
Samuel R. Scarano[1] and Jennifer B. Wagner[1,2]*

[1]Department of Psychology, College of Staten Island, City University of New York, New York, NY, United States, [2]Department of Psychology, The Graduate Center, City University of New York, New York, NY, United States, [3]Mrs. T.H. Chan Division of Occupational Science and Occupational Therapy, University of Southern California, Los Angeles, CA, United States, [4]Albert Einstein College of Medicine, The Bronx, NY, United States

The pupil light reflex (PLR), a marker of neuronal response to light, is a well-studied index of autonomic functioning. Studies have found that autistic children and adults have slower and weaker PLR responses compared to non-autistic peers, suggesting lower autonomic control. Altered autonomic control has also been associated with increased sensory difficulties in autistic children. With autistic traits varying in the general population, recent studies have begun to examine similar questions in non-autistic individuals. The current study looked at the PLR in relation to individual differences in autistic traits in non-autistic children and adults, asking how differences in the PLR could lead to variation in autistic traits, and how this might change across development. Children and adults completed a PLR task as a measure of sensitivity to light and autonomic response. Results showed that, in adults, increased levels of restricted and repetitive behaviors (RRB) were associated with a weaker and slower PLR. However, in children, PLR responses were not associated with autistic traits. Differences in PLR were also found across age groups, with adults showing smaller baseline pupil diameter and stronger PLR constriction as compared with children. The current study expanded on past work to examine the PLR and autistic traits in non-autistic children and adults, and the relevance of these findings to sensory processing difficulties is discussed. Future studies should continue to examine the neural pathways that might underlie the links between sensory processing and challenging behaviors.

KEYWORDS

pupil light reflex, pupillometry, sensory sensitivity, broader autism phenotype, restricted and repetitive behaviors

Introduction

Autism spectrum disorder is a neurodevelopmental condition characterized by social interaction and communication difficulties and restricted interests and repetitive behaviors (RRB). RRB can be displayed by stereotyped or repetitive motor behaviors, focused areas of interest, insistence on sameness, and by hyper- or hypo-responsivity to sensory input (American Psychiatric Association, 2013).

A growing body of research has asked what biological mechanisms might underlie the difficulties seen in autistic individuals, including differences in cerebral activity (e.g., Eack et al., 2017; Wolff et al., 2017; Abbott et al., 2018; Jung et al., 2019; McKinnon et al., 2019; Sato and Uono, 2019; Ecker et al., 2022) and genetic factors (e.g., Cantor et al., 2018; Ramaswami and Geschwind, 2018; Waye and Cheng, 2018; Wiśniowiecka-Kowalnik and Nowakowska, 2019; Yousaf et al., 2020; Warrier et al., 2022). Another potential factor that has been examined is the autonomic nervous system (ANS), which regulates involuntary processes in the human body, such as breathing and heart rate (e.g., Iaizzo and Fitzgerald, 2015). The ANS includes two primary branches, the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS), which work cooperatively to regulate internal processes according to conditions both inside and outside of the body. The SNS prepares the body for intense physical activity as a response to a stressful event ("fight or flight" responses), while the PNS helps to maintain homeostasis during periods of rest and recuperation ("rest and digest" responses).

One common measure used to study autonomic activity is pupillometry, which assesses pupil diameter at baseline or in response to a stimulus (Beatty and Lucero-Wagoner, 2000). The primary factor that influences pupil diameter is changes in illumination, and pupil constriction or dilation are directly linked to the amount of light entering the eye. Pupil responses can reflect the interaction and balance between the sympathetic and parasympathetic branches working together to regulate pupil size at any given time (Goldwater, 1972). For example, an increase in pupil diameter, or pupil dilation, can be a result of either an increase in SNS activity or a decrease in PNS activity (Steinhauer et al., 2004). Therefore, measures of pupillary responses often indicate general autonomic activity.

Researchers have discussed indicators of subcortical activity in relation to pupillary responses (e.g., Bast et al., 2018, 2021). For example, studies have linked arousal levels, as observed by pupil constriction and dilation, to brain activity through two paths. One suggested path to changes in pupil diameter goes through the locus coeruleus and links arousal levels with cognitive and behavioral flexibility (for a review, see Poe et al., 2020). A second path goes through the superior colliculus, which is linked to attention shifting and regulating stress-induced responses, and can also underlie cognition- and behavior-related changes in pupil diameter (for a review, see Wang and Munoz, 2015). Both paths are related to activation of the PNS and SNS (Hall and Chilcott, 2018).

The pupil light reflex (PLR), which refers to changes in pupil diameter in response to a quick flash of light, is a reliable marker of autonomic function that is regularly used in clinical settings to assess neurological processes (e.g., Cocker et al., 2005), including intensive care units (e.g., Bower et al., 2021). In addition to clinical settings, the PLR is also used in non-clinical research settings (e.g., Bremner, 1999; Beatty and Lucero-Wagoner, 2000). PLR responses have been described in terms of three phases, with the initial phase of rapid constriction in response to light controlled primarily by PNS activity, the second phase characterized by a rapid dilation controlled by both the PNS and the SNS, and the third phase characterized by a slower dilation that is mainly controlled by the SNS (e.g., Wang et al., 2016). Reduced PNS responding was found to correspond to a less robust PLR in this first phase, including smaller constriction amplitude and slower latency to constrict (Levy et al., 1992; Wang et al., 2016).

Various aspects of pupillometry have been studied in autistic individuals. Measures of the initial phase of the PLR have been consistently found to differ between autistic and non-autistic individuals across numerous studies, with slower and less pronounced PLR in autism (e.g., Fan et al., 2009; Daluwatte et al., 2013, 2015; Dinalankara et al., 2017; Lynch, 2018), suggesting reduced parasympathetic activity. These diminished PLR responses were found to also correlate with more sensory processing difficulties in autistic children (Daluwatte et al., 2015). Interestingly, infants at increased likelihood for autism (by virtue of an older autistic sibling) show a *stronger* PLR response by the age of 10 months (Nyström et al., 2015), and stronger PLR responses predicted greater autism symptomology at age 3 years (Nyström et al., 2018), suggesting changes in how the PLR might relate to autism and autistic traits across development.

Results with other pupillary measures have been mixed. For example, while some studies report differences between autistic and non-autistic individuals in both baseline pupil diameter (e.g., Anderson and Colombo, 2009; Martineau et al., 2011) and task-related pupil responses (e.g., Falck-Ytter, 2008; Blaser et al., 2014; Polzer et al., 2022), other studies have found no differences (e.g., Nuske et al., 2014, 2015; Laeng et al., 2018; for a review, see de Vries et al., 2021). PLR, baseline, and task-related pupil measures have all been discussed in terms of ANS contributions (e.g., Bradley et al., 2008; Anderson and Colombo, 2009; Wang et al., 2016), but the latter two measures have also been the focus of research studying the locus coeruleus–norepinephrine (LC-NE) system, which is located in the brainstem and has roles in cognitive processes such as attention shifting and in regulating sensory processing and sympathetic activity (for a review see Steinhauer et al., 2004). More work is needed to better understand why autonomic activity and subcortical routes might relate to different traits and behaviors.

Recently, studies have examined individual differences in autistic traits in non-autistic populations, which is part of a broader autism phenotype (BAP) approach. The BAP generally refers to autistic characteristics that are seen in varying degrees across autistic individuals and their relatives, as well as non-autistic individuals (Pickles et al., 2000). Studies have examined associations between task-induced pupil responses and autistic traits in non-autistic children and adults (e.g., DiCriscio and Troiani, 2017; Turi et al., 2018; DiCriscio et al., 2019). For example, in a combined sample of autistic and non-autistic children, DiCriscio and Troiani (2017) found that smaller changes in pupil size during pupil adaptation to light were associated with more social-communicative difficulties. Additionally, adults with more autistic traits showed differential patterns of pupil response during visual perception tasks, such as increased dilation of the pupil (DiCriscio et al., 2019). Together, these studies show that pupillary autonomic markers in children and adults can also reflect individual differences that might relate to the BAP.

The objective of the present study was to expand on past BAP work to further investigate the relationship between parasympathetic activity, using PLR measures, and autistic traits in a non-autistic sample including both children and adults. Based on work with autistic individuals (e.g., Fan et al., 2009), it was hypothesized that increased autistic traits would be associated with reduced PNS activity (i.e., weaker and slower PLR responses). Additionally, the current study aimed to examine whether there are differences in pupil response patterns between children and adults. Work by Daluwatte et al. (2012) found weaker PLR responses in children younger than 8 years old, so it was anticipated that children will show weaker PLR responses than adults.

Methods

Participants

Participants included 65 non-autistic children (M_{age} = 6.20 years, *SD* = 2.68; Range: 2 to 12 years; 33 male, 32 female) and 77 non-autistic adults (M_{age} = 20.34, *SD* = 4.67; Range: 18 to 46 years; 44 male, 32 female, 1 transmale). Children were recruited through in-person recruitment events, targeted mailings, and emails to families in the New York City and New Jersey area. Adult participants were college students in an introductory psychology course who had the opportunity to participate for course credit. For adult participants, informed consent was completed prior to the study, and for children, caregivers completed informed consent. All procedures were approved by the Institutional Review Board of the College of Staten Island, City University of New York.

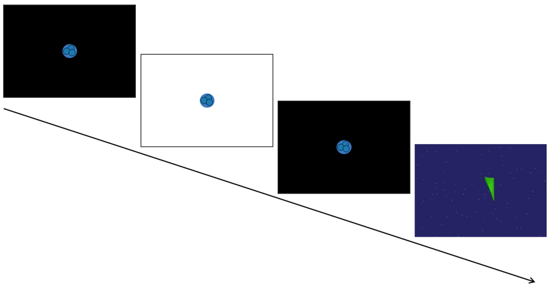


FIGURE 1
Schematic of the pupil light reflex (PLR) task. A trial consisted of a 120 ms white screen presented in between two black screens to induce the PLR response. This was followed by a brief video of moving shapes to avoid retinal saturation. [Paradigm adapted from Nyström et al. (2015)].

Procedure

A SensoMotoric Instruments (SMI) RED eye-tracking system was used to measure gaze position and pupil size at 120Hz using iView software. Pupil diameter from both eyes was collected from an average distance of 65 cm from a 22″ widescreen monitor. A 5-point calibration sequence and 4-point validation was used at the start to confirm appropriate positioning and successful tracking. Following calibration, the PLR task began based on the stimuli used in Nyström et al. (2015). Each trial totaled 6 seconds and consisted of a fixation animation on a black screen that initially lasted either 1.6, 2, or 2.4 s (varying to avoid anticipatory pupil responses), then the screen flashed white for 120 ms while the fixation animation remained on the screen, and finally the black screen with the fixation animation resumed for the remainder of the trial. In between trials, an inter-trial video of moving shapes was presented for 10 s for children and for 15 s for adults to encourage saccades and prevent retinal saturation (see **Figure 1** for schematic overview). Participants were instructed to look at the screen and attend to the PLR fixation animation until it disappeared from the screen. The experiment included nine trials, and each trial was initiated only after a clear indication that the participant was looking at the screen and the eye-tracker was successfully tracking their eye gaze. If the experimenter counted less than six potentially usable trials out of the initial nine (i.e., with attention allocated to the center of the screen before, during, and after the flash), the task was repeated and nine additional trials were presented.

Assessment of autistic traits

Autistic traits were assessed using the Social Responsiveness Scale, Second Edition (SRS-2; Constantino and Gruber, 2012), a 65-item questionnaire measure designed to examine characteristics associated with autism that has been adopted by recent studies to examine variation in these traits in the general population (e.g., DiCriscio and Troiani, 2017). Adults completed the self-report Adult Form, and caregivers completed the Preschool Form (up to 4 years) or School-Age Form (4 years and older) about their child. T-scores were calculated for SRS-2 Total score, as well as for the Social Communication and Interaction (SCI) composite and RRB subscale. Higher SRS-2 scores are associated with increased levels of autistic traits.

Data processing and analysis

Custom Python scripts were used to process the PLR dilation time series to identify PLR metrics for each trial. There were two initial inclusion criteria used for each eye for each trial: (1) no more than 100 ms of pupil data was missing during the first 1500 ms after the flash (e.g., due to blinks) and (2) valid pupil data was required at the time of the flash. Based on approaches taken in past PLR work (e.g., Fan et al., 2009; Nyström et al., 2015), pupil diameter for each eye for included trials was processed using a degree-2 Savitzky-Golay filter with a window of 11 samples to yield smoothed diameter and acceleration series, which were then further smoothed using a Gaussian convolution with a standard deviation of 5 samples. A final set of criteria were used to ensure that the resulting data accurately reflected the PLR curve: (1) the point of greatest minimum amplitude was reached within 1500 ms of the flash, (2) the point of greatest negative velocity was reached within 750 ms of the flash, and (3) the point of greatest negative acceleration was reached within 500 ms of the flash (see **Supplementary Table 1** for number of eye trials excluded at each stage of processing for children and adults).

Based on past findings with infants, children, and adults, pupil measures calculated during the PLR task included (a) *baseline pupil diameter* (A_0; e.g., Anderson and Colombo, 2009), (b) *relative constriction amplitude*, calculated from A_0 and A_m (minimum diameter) as $(A_0^2 - A_m^2)/A_0^2$ (e.g., Fan et al., 2009), (c) *absolute constriction amplitude*, A_0-A_m (e.g., DiCriscio and Troiani, 2017); and (d) *median constriction latency*, calculated as median latency to reach maximum negative acceleration (e.g., Nyström et al., 2015). Of these four measures, it should be noted that relative constriction amplitude and constriction latency have been most consistently found to reflect PNS activity (Wang et al., 2016). When clean data was available for both eyes on a given trial, PLR variables were averaged across both eyes, and then PLR metrics were averaged across usable trials for each participant.

PLR analyses focused on participants with four or more valid trials (M_{trials} = 7.08, *SD* = 1.73, range: 4-11; e.g., Nyström et al., 2015), leading to the exclusion of six children and eight adults. Three additional adults were excluded because they were age outliers (see Statistical analysis for more information). The final included sample therefore included 59 children (M_{age} = 6.36 years, *SD* = 2.72 years, age range: 2-12 years) and 66 adults (M_{age} = 19.64 years, *SD* = 2.04 years, age range: 18-28 years). For an illustration of the average PLR response over time for adults and children see **Supplementary Figure 1**, and for histograms illustrating distributions of the PLR measures see **Supplementary Figure 2**.

Statistical analysis

The primary analyses included (1) a series of correlations to examine associations between PLR measures and autistic traits for each group, based on the SRS-2, and (2) a series of independent samples t-tests to examine developmental differences in the PLR between children and adults. Prior work across childhood (e.g., Daluwatte et al., 2012) and adulthood (e.g., Telek et al., 2018) has found age to be a significant factor in pupillary responses to light. Because of these past findings, and due to the wide age ranges for both groups, an age outlier check was conducted within each group. Participants who fell more than 3 SDs above or below the age mean were excluded from subsequent analyses. This resulted in the exclusion of three adult participants (aged 37 to 46 years; see Telek et al., 2018 for discussion of adult age-related differences).

With age outliers removed, a series of preliminary correlations were run to examine the relationship between age and PLR measures within each sample. Results showed that in children, age was positively associated with baseline pupil diameter ($r(53) = 0.31$, $p = 0.020$), suggesting that older children have greater pupil diameter at baseline. In adults, no associations were found between age and PLR measures (*ps* > 0.22). In subsequent correlational analyses, because age was associated with PLR measures in children, partial correlations controlling for age were used for the child sample, while standard bivariate correlations were used for adults.

Results

Correlational analyses

Relations among pupil measures

An initial set of correlations examined relations among the four pupil response measures, using a Bonferroni correction accounting for six comparisons for each age group (critical $p = 0.05 / 6 = 0.0083$). Analyses included partial correlations accounting for age for children, and bivariate correlations for adults. At both ages, several variables were significantly correlated with each other (see **Supplementary Tables 2, 3** for correlation tables): PLR absolute constriction amplitude was positively associated with baseline pupil diameter and with PLR relative constriction amplitude (*ps* < 0.001). Further, in the adult sample only, PLR constriction latency was negatively associated

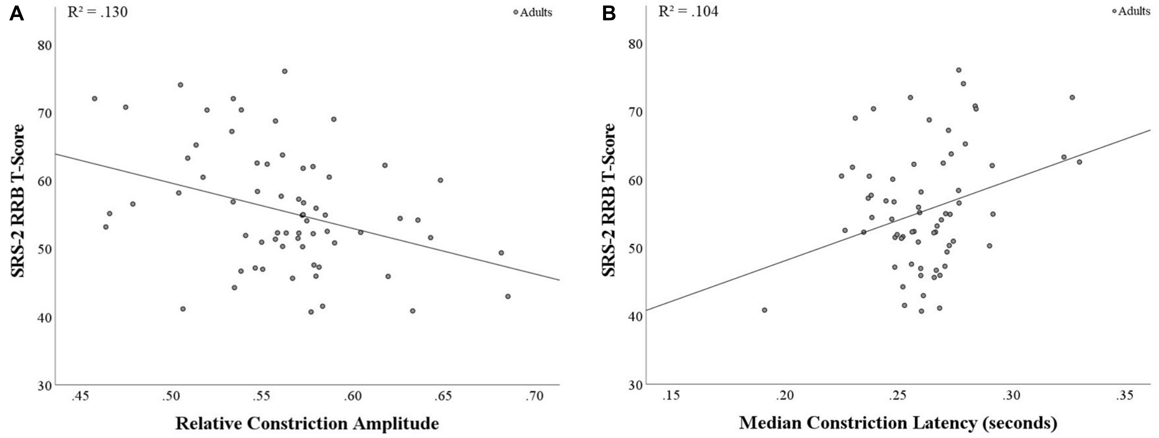


FIGURE 2
Correlations between PLR metrics and SRS-2 RRB score in adults. **(A)** A significant negative correlation was found between relative constriction amplitude and RRB ($p = 0.003$), with stronger PLR response associated with lower RRB scores. **(B)** A significant positive correlation was found between median latency and RRB ($p = 0.008$), with faster time to the point of maximum negative acceleration (i.e., shorter PLR response) associated with lower RRB scores.

with PLR relative constriction amplitude ($p = 0.002$). No other results held after the corrected p-value (see Supplementary Tables 2, 3).

Relations between autistic traits and pupil measures

The primary correlational analyses examined relations between autistic traits and pupil responses in children and adults, using a Bonferroni correction taking into account associations between SRS-2 scores and the four different pupil measures (critical $p = 0.05 / 4 = 0.0125$). For children, partial correlations were used, controlling for age, and for adults, bivariate correlations were used (see Supplementary Tables 4, 5 for the full results).

Children. After controlling for age, findings showed that RRB was negatively associated with baseline pupil diameter ($r(50) = -0.32$, $p = 0.022$), however, this finding did not survive the corrected p-value. Non-significant trends were also found that suggested greater absolute constriction amplitude was marginally related to lower levels of autistic traits overall, as well as SCI specifically ($rs > -0.25$, $ps < 0.10$; see Supplementary Table 4 for full results).

Adults. Bivariate correlations showed a significant negative correlation between SRS-2 Total score and relative constriction amplitude ($r(64) = -0.28$, $p = 0.024$), however this finding did not survive the corrected p-value. A non-significant trend was also found between SRS-2 Total and median latency ($r(64) = 0.21$, $p = 0.092$). No other PLR measures were significantly associated with overall level of autistic traits ($ps > 0.40$).

When examining correlations between SRS-2 SCI and RRB scores in relation to pupil measures, RRB was found to be negatively associated with relative constriction amplitude ($r(64) = -0.36$, $p = 0.003$; see Figure 2A) and positively associated with median latency ($r(64) = 0.32$, $p = 0.008$; see Figure 2B), with both findings surviving the corrected p-value. This suggests that increased levels of RRB are associated with smaller relative pupil constriction and longer latency to respond to light, indicating weaker and slower PLR. Additionally, SRS-2 SCI and relative constriction amplitude were marginally associated ($r(64) = -0.21$, $p = 0.088$; see Supplementary Table 5 for full results).

Group comparisons

A series of independent-samples t-tests examined differences in pupillary responses between children and adults. A Bonferroni correction was applied, taking into account group comparisons for the four different pupil measures (critical $p = 0.05 / 4 = 0.0125$).

When comparing adults and children on the pupil measures, adults were found to have smaller baseline pupil diameter than children ($t(119) = 5.88$, $p < 0.001$, Cohen's $d = 1.07$; see Figure 3A). Additionally, adults showed greater relative constriction amplitude than children ($t(119) = 5.12$, $p < 0.001$, Cohen's $d = 0.94$; see Figure 3B), but no differences were found for median latency or absolute constriction amplitude ($ps > 0.30$; see Supplementary Figures 1, 2 for further data visualization). All results held with and without correction.

Discussion

The current study used a PLR task adapted from Nyström et al. (2015) and had two main aims: first, to examine PLR

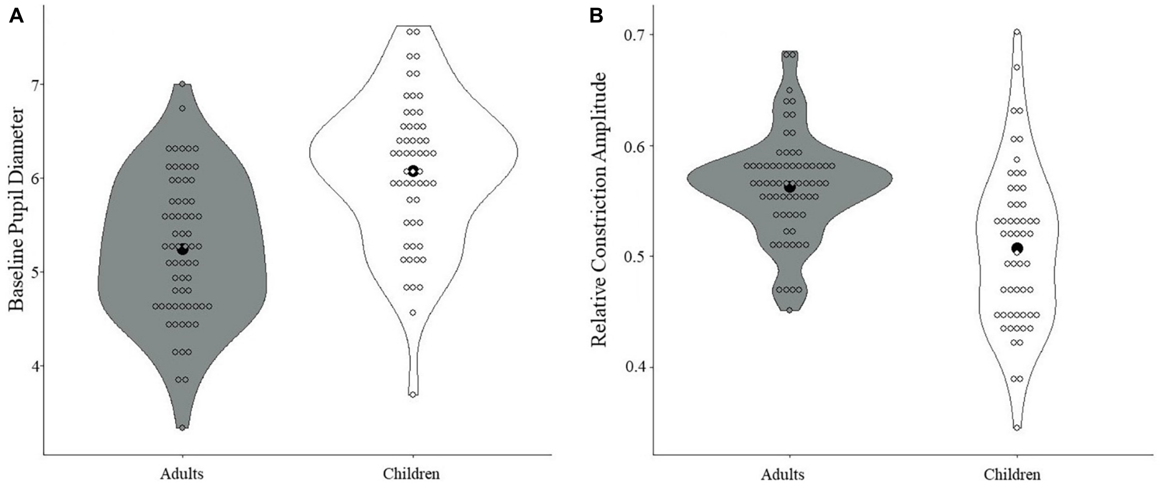


FIGURE 3
Differences in PLR measures between children and adults. **(A)** A significant difference in baseline pupil diameter was found, with smaller baseline pupil diameter in adults compared with children ($p < 0.001$). **(B)** A significant difference in relative constriction amplitude was found, with greater relative constriction amplitude in adults than in children ($p < 0.001$). Black dots denote the mean.

responses in non-autistic children and adults in relation to levels of autistic traits, and second, to examine the differences in PLR responses across age groups. Main findings showed that (1) adults with increased levels of RRB showed a less robust PLR response (i.e., slower and weaker constriction), and (2) children showed larger pupil diameter at baseline and weaker relative PLR constriction compared with adults.

In relation to autistic traits, the current study showed that in children, after controlling for age, no relations between autistic traits and pupil measures survived correction for multiple comparison. However, trends were found showing that children who have increased levels of RRB also have smaller pupil diameter at baseline. Additionally, trends were found whereby higher levels of autistic traits overall and the SCI composite were both marginally correlated with a smaller absolute amplitude change during the PLR. Although this was not significant in the current sample, these trends align with findings from DiCriscio and Troiani (2017), showing that changes in pupil size during light adaptation were associated with differences in autistic traits in children, with significant results relating to the SRS-2 SCI composite score, but not the RRB subscale (DiCriscio and Troiani, 2017). Relatedly, in work with infants with and without an older autistic sibling, PLR relative constriction amplitude at 9 to 10 months predicted autistic traits at three years old, again, with significant findings focused on overall scores and social metrics, but not RRB (Nyström et al., 2018). More work is needed to understand why PLR metrics related to PNS responding might be more predictive of social-communication measures in children as compared to RRB, which might relate more to baseline pupil measures as suggested by the current work.

In adults, after controlling for multiple comparisons, overall levels of autistic traits showed a trend towards a negative association with relative constriction. When examining the subscales of autistic traits (SCI and RRB) in relation to PLR responses, significant findings after correction indicated that adults with increased levels of RRB showed both smaller relative constriction and longer latency to reach the point of constriction onset (point of maximum negative acceleration; e.g., Fan et al., 2009), indicating weaker and slower PLR on two well-studied markers of parasympathetic control. These findings suggest that, among a non-clinical sample of adults, those with better parasympathetic control endorsed fewer RRB. Although PLR measures have not been found to relate consistently to RRB in children, research in autistic and non-autistic children using cardiac autonomic measures has also found that increased respiratory sinus arrhythmia, a measure of better parasympathetic control, was related to lower levels of RRB across both groups (Condy et al., 2017). Taken together, these studies point to a role for parasympathetic markers in predicting adaptive functioning in the domain of RRB, but suggest that pupil measures and cardiac measures may be related to RRB at different points in development. It is important to note that the findings above in children showed no association between RRB and PNS-related PLR measures, in contrast to the findings with the adult sample. Further research is needed to clarify why different PLR measures were associated with autistic traits in adults but not in children, and to elucidate the mechanisms that might underlie these non-parallel results.

Examining past work linking sensory processing and RRB allows for a better understanding of the current associations between PLR and increased RRB in adults. Past research suggests that increased RRB are associated with difficulties in sensory processing in adults (Hwang et al., 2020; Moore et al., 2021) and children (Gabriels et al., 2008; Chen et al., 2009; Boyd et al., 2010; Schulz and Stevenson, 2019), and that smaller PLR constriction is related to more atypical sensory features in autistic children (Daluwatte et al., 2015). Increased sensitivity to sensory input, such as light, can lead to experienced overstimulation that might cause distress that needs to be regulated with the first available self-regulation method, such as RRB (e.g., Militerni et al., 2002; Baker et al., 2008). Because the PLR is a neurological measure of sensitivity to light (Beatty and Lucero-Wagoner, 2000; Lynch, 2018), increased RRB in relation to the PLR might imply elevated sensitivity to sensory stimuli in the environment. Results from the current study provide further support for the link between sensory sensitivity and RRB, as weaker PLR responses might indicate that the pupil diameter is not effectively and optimally regulating the amount of light that enters the eye, which can lead to more intense experiences with visual stimuli in the environment. With the PLR reflecting a key neural pathway in visual sensory processing, and with variation in sensory processing in autistic and non-autistic individuals relating to behavioral challenges, it could be posited that the relationship between sensory processing and some autism-related behaviors might be driven by this physiological mechanism.

In the present study, when examining overall developmental differences between children and adults, children showed increased pupil diameter at baseline in comparison to adults (for related work, see Telek et al., 2018). Additionally, children showed smaller relative constriction of the pupil in response to light, and this weaker PLR response in children also suggests weaker parasympathetic control. In a study that examined developmental trajectories of PLR responses in children and adolescents between 6 and 17 years of age, relative pupil constriction was found to increase (i.e., PLR became stronger) between the ages of 6 and 8 years, and then stabilized from ages 8 to 17 years (Daluwatte et al., 2012). In the current study, children and adults differed in relative constriction levels, but when looking at each age group separately, relative constriction was not associated with age. Although several studies have attempted to capture developmental changes in PLR metrics (Dinalankara et al., 2017; Telek et al., 2018), the specific developmental trajectory of PLR responses from infancy to adulthood in non-clinical populations is not yet clear, especially when examining it continuously. Future work should continue to examine trajectories of parasympathetic activity from infancy across development through changes in pupillary responding, exploring how these trends change across a wide age range.

The current study had several limitations. First, because the current sample did not include participants between the ages of 12 and 18 years, it was not possible to examine age as a continuous measure, limiting conclusions about the developmental trends in sensory responding seen through the PLR. This will be important to examine in future research, especially as hormonal changes associated with puberty might play a role in changes across age. A second limitation of the current study is that baseline pupil diameter was calculated during the PLR task, just before the flash occurred, and no baseline measurement outside the task was recorded. Future work should examine how differences in baseline calculation might affect age-related differences in pupillary measures, and how different baseline calculations might differ in relation to autistic traits.

Extending previous work that has found a less robust PLR response in autistic children (e.g., Fan et al., 2009) and a negative association between light adaptation responses and autistic traits in a broad population of children (DiCriscio and Troiani, 2017), the current study found that young adults with weaker and slower PLR have higher levels of RRB. Altogether, this points to the PLR as a marker associated with the broader autism phenotype, as opposed to an autism diagnosis. This well-studied marker of autonomic functioning could therefore provide an important window into the study of individual differences in adaptive behavior for both autistic and non-autistic individuals.

Data availability statement

The de-identified data that support the findings of this study are available on request from the corresponding author JBW, jen.wagner@alum.mit.edu. The data are not publicly available due to privacy restrictions.

Ethics statement

The present study involving human participants was reviewed and approved by the Institutional Review Board of the College of Staten Island, City University of New York (Protocol #570016). Adult participants provided written informed consent to participate in this study; for child participants, written informed consent was provided by the child's legal guardian.

Author contributions

SS-E contributed to the design of the study, data collection, organization, processing, statistical analysis, and wrote the initial draft of the current manuscript. AL contributed to data collection and the writing of portions of a previous version of the manuscript. SRS developed Python scripts for pupillometry data preparation and analysis. JBW contributed to the conception, design of the study and approach to analysis, and contributed to editing of all sections of the manuscript.

All authors read and approved the final submitted version of the manuscript.

Funding

This research was made possible, in part, by grants from NIMH (R15 MH112090) and PSC-CUNY (Awards 69677-00-47 and 64484-00-52).

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2022.1052604/full#supplementary-material

References

Abbott, A., Linke, A., Nair, A., Jahedi, A., Alba, L., Keown, C., et al. (2018). Repetitive behaviors in autism are linked to imbalance of corticostriatal connectivity: A functional connectivity MRI study. *Soc. Cogn. Affect. Neurosci.* 13, 32–42. doi: 10.1093/scan/nsx129

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders fifth Edition*. Arlington: American Psychiatric Association, 991. doi: 10.1176/appi.books.9780890425596

Anderson, C., and Colombo, J. (2009). Larger tonic pupil size in young children with autism spectrum disorder. *Dev. Psychobiol.* 51, 207–211. doi: 10.1002/dev.20352

Baker, A., Lane, A., Angley, M., and Young, R. (2008). The relationship between sensory processing patterns and behavioural responsiveness in autistic disorder: A pilot study. *J. Autism Dev. Disord.* 38, 867–875. doi: 10.1007/s10803-007-0459-0

Bast, N., Boxhoorn, S., Supér, H., Helfer, B., Polzer, L., Klein, C., et al. (2021). Atypical arousal regulation in children with autism but not with attention-deficit/hyperactivity disorder as indicated by pupillometric measures of locus coeruleus activity. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* S2451-9022, 117–118. doi: 10.1016/j.bpsc.2021.04.010

Bast, N., Poustka, L., and Freitag, C. (2018). The locus coeruleus–norepinephrine system as pacemaker of attention – a developmental mechanism of derailed attentional function in autism spectrum disorder. *Eur. J. Neurosci.* 47, 115–125. doi: 10.1111/ejn.13795

Beatty, J., and Lucero-Wagoner, B. (2000). "The pupillary system," in *Handbook of psychophysiology*, eds J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson (Cambridge: Cambridge University Press), 142–162.

Blaser, E., Eglington, L., Carter, A., and Kaldy, Z. (2014). Pupillometry reveals a mechanism for the autism spectrum disorder (asd) advantage in visual tasks. *Sci. Rep.* 4:4301. doi: 10.1038/srep04301

Bower, M., Sweidan, A., Xu, J., Stern-Neze, S., Yu, W., and Groysman, L. (2021). Quantitative pupillometry in the intensive care unit. *J. Intens. Care Med.* 36, 383–391. doi: 10.1177/0885066619881124

Boyd, B., Baranek, G., Sideris, J., Poe, M., Watson, L., Patten, E., et al. (2010). Sensory features and repetitive behaviors in children with autism and developmental delays. *Autism Res.* 3, 78–87. doi: 10.1002/aur.124

Bradley, M., Miccoli, L., Escrig, M., and Lang, P. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* 45, 602–607. doi: 10.1111/j.1469-8986.2008.00654.x

Bremner, F. (1999). in *The Pupil: Anatomy, physiology, and clinical applications*, Vol. 2001, ed. E. Loewenfeld (Oxford: Butterworth-Heinemann), 1881–1883. doi: 10.1093/brain/124.9.1881

Cantor, R., Navarro, L., Won, H., Walker, R., Lowe, J., and Geschwind, D. H. (2018). ASD restricted and repetitive behaviors associated at 17q21.33: Genes prioritized by expression in fetal brains. *Mol. Psychiatry* 23, 993–1000. doi: 10.1038/mp.2017.114

Chen, Y., Rodgers, J., and McConachie, H. (2009). Restricted and repetitive behaviours, sensory processing and cognitive style in children with autism spectrum disorders. *J. Autism Dev. Disord.* 39, 635–642. doi: 10.1007/s10803-008-0663-6

Cocker, K., Fielder, A., Moseley, M., and Edwards, A. (2005). Measurements of pupillary responses to light in term and preterm infants. *Neuro Ophthalmol.* 29, 95–101. doi: 10.1080/01658100590958274

Condy, E., Scarpa, A., and Friedman, B. (2017). Respiratory sinus arrhythmia predicts restricted repetitive behavior severity. *J. Autism Dev. Disord.* 47, 2795–2804. doi: 10.1007/s10803-017-3193-2

Constantino, J. N., and Gruber, C. P. (2012). *Social responsiveness scale: SRS-2*. Torrance, CA: Western Psychological Services.

Daluwatte, C., Miles, J., Christ, S., Beversdorf, D., Lofgreen, A., Berliner, N., et al. (2012). "Age-dependent pupillary light reflex parameters in children," in *In Proceedinds of the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, (Piscataway), 3776–3779. doi: 10.1109/EMBC.2012.6346789

Daluwatte, C., Miles, J., Christ, S., Beversdorf, D., Takahashi, T., and Yao, G. (2013). Atypical pupillary light reflex and heart rate variability in children with autism spectrum disorder. *J. Autism Dev. Disord.* 43, 1910–1925. doi: 10.1007/s10803-012-1741-3

Daluwatte, C., Miles, J., Sun, J., and Yao, G. (2015). Association between pupillary light reflex and sensory behaviors in children with autism spectrum disorders. *Res. Dev. Disabil.* 37, 209–215. doi: 10.1016/j.ridd.2014.11.019

de Vries, L., Fouquaet, I., Boets, B., Naulaers, G., and Steyaert, J. (2021). Autism spectrum disorder and pupillometry: A systematic review and meta-analysis. *Neurosci. Biobehav. Rev.* 120, 479–508. doi: 10.1016/j.neubiorev.2020.09.032

DiCriscio, A., and Troiani, V. (2017). Pupil adaptation corresponds to quantitative measures of autism traits in children. *Sci. Rep.* 7:6476. doi: 10.1038/s41598-017-06829-1

DiCriscio, A., Hu, Y., and Troiani, V. (2019). Brief report: Pupillometry. Visual perception, and ASD Features in a task-switching paradigm. *J. Autism Dev. Disord.* 49, 5086–5099. doi: 10.1007/s10803-019-04213-8

Dinalankara, D., Miles, J., Nicole Takahashi, T., and Yao, G. (2017). Atypical pupillary light reflex in 2–6-year-old children with autism spectrum disorders. *Autism Res.* 10, 829–838. doi: 10.1002/aur.1745

Eack, S., Wojtalik, J., Keshavan, M., and Minshew, N. (2017). Social-cognitive brain function and connectivity during visual perspective-taking in autism and schizophrenia. *Schizophr. Res.* 183, 102–109. doi: 10.1016/j.schres.2017.03.009

Ecker, C., Pretzsch, C., Bletsch, A., Mann, C., Schaefer, T., Ambrosino, S., et al. (2022). Interindividual differences in cortical thickness and their genomic underpinnings in autism spectrum disorder. *Am. J. Psychiatry* 179, 242–254. doi: 10.1176/appi.ajp.2021.20050630

Falck-Ytter, T. (2008). Face inversion effects in autism: A combined looking time and pupillometric study. *Autism Res. Off. J. Int. Soc. Autism Res.* 1, 297–306. doi: 10.1002/aur.45

Fan, X., Miles, J., Takahashi, N., and Yao, G. (2009). Abnormal transient pupillary light reflex in individuals with autism spectrum disorders. *J. Autism Dev. Disord.* 39, 1499–1508. doi: 10.1007/s10803-009-0767-7

Gabriels, R., Agnew, J., Miller, L., Gralla, J., Pan, Z., Goldson, E., et al. (2008). Is there a relationship between restricted, repetitive, stereotyped behaviors and interests and abnormal sensory response in children with autism spectrum disorders? *Res. Autism Spectr. Disord.* 2, 660–670. doi: 10.1016/j.rasd.2008.02.002

Goldwater, B. (1972). Psychological significance of pupillary movements. *Psychol. Bull.* 77, 340–355. doi: 10.1037/h0032456

Hall, C., and Chilcott, R. (2018). Eyeing up the future of the pupillary light reflex in neurodiagnostics. *Diagnostics* 8:19. doi: 10.3390/diagnostics8010019

Hwang, Y., Arnold, S., Srasuebkul, P., and Trollor, J. (2020). Understanding anxiety in adults on the autism spectrum: An investigation of its relationship with intolerance of uncertainty, sensory sensitivities and repetitive behaviours. *Autism* 24, 411–422. doi: 10.1177/1362361319868907

Iaizzo, P., and Fitzgerald, K. (2015). "Autonomic nervous system," in *Handbook of cardiac anatomy, physiology, and devices [Internet]*, ed. P. Iaizzo (Cham: Springer International Publishing), 235–250. doi: 10.1007/978-3-319-19464-6_14

Jung, M., Tu, Y., Lang, C., Ortiz, A., Park, J., Jorgenson, K., et al. (2019). Decreased structural connectivity and resting-state brain activity in the lateral occipital cortex is associated with social communication deficits in boys with autism spectrum disorder. *NeuroImage* 190, 205–212. doi: 10.1016/j.neuroimage.2017.09.031

Laeng, B., Færevaag, F., Tanggaard, S., and von Tetzchner, S. (2018). Pupillary responses to illusions of brightness in autism spectrum disorder. *Percept* 9:2041669518771716. doi: 10.1177/2041669518771716

Levy, D., Rowley, D., and Abraham, R. (1992). Portable infrared pupillometry using Pupilscan: Relation to somatic and autonomic nerve function in diabetes mellitus. *Clin. Auton. Res.* 2, 335–341. doi: 10.1007/BF01824304

Lynch, G. (2018). Using pupillometry to assess the atypical pupillary light reflex and LC-NE System in ASD. *Behav. Sci.* 8:108. doi: 10.3390/bs8110108

Martineau, J., Hernandez, N., Hiebel, L., Roché, L., Metzger, A., and Bonnet-Brilhault, F. (2011). Can pupil size and pupil responses during visual scanning contribute to the diagnosis of autism spectrum disorder in children? *J. Psychiatr. Res.* 45, 1077–1082. doi: 10.1016/j.jpsychires.2011.01.008

McKinnon, C., Eggebrecht, A., Todorov, A., Wolff, J., Elison, J., Adams, C., et al. (2019). Restricted and repetitive behavior and brain functional connectivity in infants at risk for developing autism spectrum disorder. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 4, 50–61. doi: 10.1016/j.bpsc.2018.09.008

Militerni, R., Bravaccio, C., Falco, C., Fico, C., and Palermo, M. (2002). Repetitive behaviors in autistic disorder. *Eur. Child Adolesc. Psychiatry* 11, 210–218. doi: 10.1007/s00787-002-0279-x

Moore, H., Brice, S., Powell, L., Ingham, B., Freeston, M., Parr, J., et al. (2021). The mediating effects of alexithymia, intolerance of uncertainty, and anxiety on the relationship between sensory processing differences and restricted and repetitive behaviours in autistic adults. *J. Autism Dev. Disord.* 52, 4384–4396. doi: 10.1007/s10803-021-05312-1

Nuske, H., Vivanti, G., and Dissanayake, C. (2014). Brief report: Evidence for normative resting-state physiology in autism. *J. Autism Dev. Disord.* 44, 2057–2063. doi: 10.1007/s10803-014-2068-z

Nuske, H., Vivanti, G., and Dissanayake, C. (2015). No evidence of emotional dysregulation or aversion to mutual gaze in preschoolers with autism spectrum disorder: An eye-tracking pupillometry study. *J. Autism Dev. Disord.* 45, 3433–3445. doi: 10.1007/s10803-015-2479-5

Nyström, P., Gliga, T., Jobs, E., Gredebäck, G., Charman, T., Johnson, M., et al. (2018). Enhanced pupillary light reflex in infancy is associated with autism diagnosis in toddlerhood. *Nat. Commun.* 9, 1–5. doi: 10.1038/s41467-018-03985-4

Nyström, P., Gredebäck, G., Bölte, S., and Falck-Ytter, T. (2015). Hypersensitive pupillary light reflex in infants at risk for autism. *Mol. Autism* 6, 1–6. doi: 10.1186/s13229-015-0011-6

Pickles, A., Starr, E., Kazak, S., Bolton, P., Papanikolaou, K., Bailey, A., et al. (2000). Variable expression of the autism broader phenotype: Findings from extended pedigrees. *J. Child Psychol. Psychiatry* 41, 491–502. doi: 10.1111/1469-7610.00634

Poe, G., Foote, S., Eschenko, O., Johansen, J., Bouret, S., Aston-Jones, G., et al. (2020). Locus coeruleus: A new look at the blue spot. *Nat. Rev. Neurosci.* 21, 644–659. doi: 10.1038/s41583-020-0360-9

Polzer, L., Freitag, C., and Bast, N. (2022). Pupillometric measures of altered stimulus-evoked locus coeruleus-norepinephrine activity explain attenuated social attention in preschoolers with autism spectrum disorder. *Autism Res.* 15, 2167–2180. doi: 10.1002/aur.2818

Ramaswami, G., and Geschwind, D. (2018). *Genetics of autism spectrum disorder. In: Handbook of Clinical Neurology*. Amsterdam: Elsevier B.V, 321–329. doi: 10.1016/B978-0-444-63233-3.00021-X

Sato, W., and Uono, S. (2019). The atypical social brain network in autism: Advances in structural and functional MRI studies. *Curr. Opin. Neurol.* 32, 617–621. doi: 10.1097/WCO.0000000000000713

Schulz, S., and Stevenson, R. (2019). Sensory hypersensitivity predicts repetitive behaviours in autistic and typically-developing children. *Autism* 23, 1028–1041. doi: 10.1177/1362361318774559

Steinhauer, S., Siegle, G., Condray, R., and Pless, M. (2004). Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *Int. J. Psychophysiol.* 52, 77–86. doi: 10.1016/j.ijpsycho.2003.12.005

Telek, H., Erdol, H., and Turk, A. (2018). The effects of age on pupil diameter at different light amplitudes. *Beyoglu Eye J.* 3, 80–85. doi: 10.14744/bej.2018.43534

Turi, M., Burr, D., and Binda, P. (2018). Pupillometry reveals perceptual differences that are tightly linked to autistic traits in typical adults. *eLife* 7:e32399. doi: 10.7554/eLife.32399

Wang, C., and Munoz, D. P. (2015). A circuit for pupil orienting responses: Implications for cognitive modulation of pupil size. *Curr. Opin. Neurobiol.* 33, 134–140. doi: 10.1016/j.conb.2015.03.018

Wang, Y., Zekveld, A., Naylor, G., Ohlenforst, B., Jansma, E., Lorens, A., et al. (2016). Parasympathetic nervous system dysfunction, as identified by pupil light reflex, and its possible connection to hearing impairment. *PLoS One* 11:e0153566. doi: 10.1371/journal.pone.0153566

Warrier, V., Zhang, X., Reed, P., Havdahl, A., Moore, T., Cliquet, F., et al. (2022). Genetic correlates of phenotypic heterogeneity in autism. *Nat. Genet.* 54, 1293–1304. doi: 10.1038/s41588-022-01072-5

Waye, M., and Cheng, H. (2018). Genetics and epigenetics of autism: A Review. *Psychiatry Clin. Neurosci.* 72, 228–244. doi: 10.1111/pcn.12606

Wiśniowiecka-Kowalnik, B., and Nowakowska, B. (2019). Genetics and epigenetics of autism spectrum disorder—current evidence in the field. *J. Appl. Genet.* 60, 37–47. doi: 10.1007/s13353-018-00480-w

Wolff, J., Swanson, M., Elison, J., Gerig, G., Pruett, J., Styner, M., et al. (2017). Neural circuitry at age 6 months associated with later repetitive behavior and sensory responsiveness in autism. *Mol. Autism* 8, 1–12.

Yousaf, A., Waltes, R., Haslinger, D., Klauck, S., Duketis, E., Sachse, M., et al. (2020). Quantitative genome-wide association study of six phenotypic subdomains identifies novel genome-wide significant variants in autism spectrum disorder. *Transl. Psychiatry* 10, 1–11. doi: 10.1038/s41398-020-00906-2

frontiers | Frontiers in Human Neuroscience

TYPE Original Research
PUBLISHED 28 February 2023
DOI 10.3389/fnhum.2022.1043475

OPEN ACCESS

EDITED BY
Alessia Tonelli,
Italian Institute of Technology (IIT), Italy

REVIEWED BY
Peter Gerhardstein,
Binghamton University, United States
Mauro Manassi,
University of Aberdeen, United Kingdom

*CORRESPONDENCE
Alexandra List
✉ alist@hamilton.edu

SPECIALTY SECTION
This article was submitted to
Sensory Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 13 September 2022
ACCEPTED 14 December 2022
PUBLISHED 28 February 2023

CITATION
List A (2023) Global and local priming in a multi-modal context.
Front. Hum. Neurosci. 16:1043475.
doi: 10.3389/fnhum.2022.1043475

Global and local priming in a multi-modal context

Alexandra List*

Department of Psychology and Neuroscience Program, Hamilton College, Clinton, NY, United States

Perceptual information can be processed at many different scales, from featural details to entire scenes. Attentional selection of different scales has been studied using hierarchical stimuli, with research elucidating a variety of biases in local and global attentional selection (due to, e.g., stimulus properties, brain injury, and experience). In this study, the emphasis is on biases produced through recent experience, or level-specific priming effects, which have been demonstrated within both the visual and auditory modalities. Namely, when individuals attend to local information, they are subsequently biased to attend locally (and similarly so with global attention). Here, these level-specific priming effects are investigated in a multi-modal context to determine whether cross-modal interactions occur between visual and auditory modalities during hierarchical processing. Specifically, the study addresses if attentional selection of local or global information in the visual modality subsequently biases auditory attentional selection to that level, and vice versa (i.e., level-priming). Though expected identity priming effects emerged in the study, no cross-modal *level-priming* effects manifested. Furthermore, the multi-modal context eliminated the well-established within-modality level-specific priming effects. Thus, though the study does reveal a multi-modal effect, it was not a level-based effect. Instead, paradoxically, the multi-modal context eliminated attentional scope biases (i.e., level-priming) within uni-modal transitions. In other words, when visual and auditory information are equally likely require attention, no persistence emerges for processing local or global information over time, even within a single modality.

KEYWORDS

global, local, vision, audition, cross-modal processing, priming

1. Introduction

Our perceptual environment can be appreciated at many different scales. Visually, individuals can attend to an entire scene, objects within a scene, parts of objects and even object surface and textural qualities. The ability to adjust attentional scope has been studied using hierarchical figures in which local elements and global configurations can be independently manipulated (e.g., Navon, 1977; Kinchla and Wolfe, 1979). For example, local Es can be arranged to create a global H, and a person can flexibly identify the information at either level (local or global). In his influential report, Navon (1977) argued that participants show global precedence, wherein global information processing is prioritized over local. However, various later studies have shown that attentional scope biases are more flexible, and shift depending on stimulus parameters. For example, attentional biases to local or global information can vary depending on the absolute size of the hierarchical stimuli (local biases are more likely with larger stimuli; Kinchla and Wolfe, 1979; Lamb and Robertson, 1990), stimulus eccentricity in the visual field (global biases are more likely with more peripheral stimuli; Lamb and Robertson, 1988), and the density (global biases are more likely with denser local elements; Martin, 1979) or number (local biases are more likely with fewer local elements; Kimchi and Palmer, 1982) of local elements.

Attentional biases to local and global information has also been shown to depend on interhemispheric processing balance (see Ivry and Robertson, 1998 for a broad survey), perhaps most convincingly from studies of brain injured individuals. Specifically, right-hemisphere injuries produce a local bias and left-hemisphere injuries produce a global bias (e.g., Delis et al., 1986; Lamb et al., 1988, 1989; Robertson et al., 1988). Rafal and Robertson (1995) even argued that right-hemisphere local biases are likely contributors to hemi-spatial neglect, further exacerbating a rightward spatial bias by limiting patients' abilities to expand their attentional window. Indeed, Bultitude et al. (2009) showed that prism adaptation, a method more commonly used to alleviate lateralized spatial biases in hemi-spatial neglect (e.g., Rossetti et al., 1998; Bultitude and Rafal, 2010), increased global processing in individuals with right temporal-parietal brain injuries. These studies of brain injured individuals provide support for the notion that the two hemispheres contribute complementarily in controlling attentional scope.

Not only do stimulus attributes and functional inter-hemispheric balance contribute to attentional scope, but so does recent experience. In healthy individuals, how someone has deployed their attention in one moment will impact their ensuing attentional scope (e.g., Ward, 1982; Robertson, 1996; Filoteo et al., 2001; List et al., 2013). Without an incentive otherwise, when individuals attend to global information, they are subsequently biased to (again) attend to global information. Similarly, attending to local information will subsequently bias attention to local information. These effects are described as level-priming, which Robertson (1996) attributed to an *attentional persistence.* Critically, level-priming is independent of identity or response priming, because it occurs whether or not a repetition of target shape or response also occurs (Robertson, 1996; also see Filoteo et al., 2001). Furthermore, level-priming is also robust to changes of stimulus location or surface attributes (Lamb and Robertson, 1988; Robertson, 1996) or to absolute stimulus size (e.g., Kim et al., 1999). Thus, level-priming has been well-isolated from other priming effects, suggesting that the scope of attentional selection is indeed what is being primed. Robertson (1996) argued that attentional persistence only arises when hierarchical parsing is necessary, in which case the attentional selection process leaves a trace, which then biases subsequent selection.

Though most research on attention to hierarchical information has been conducted in the visual modality, various studies have shown that attention to different stimulus scopes also occurs in audition (e.g., Justus and List, 2005; Sanders and Poeppel, 2007; List and Justus, 2010; Ouimet et al., 2012). In auditory studies, as in visual ones, local stimulus patterns are repeated to create an overall global pattern. For example, in **Figure 1**, the top left hierarchical pattern represents a three-element "falling-rising" pattern repeated three times to create a global "rising-rising" pattern (imagine time elapsing on the *x*-axis and frequency on the *y*-axis, as in musical notation). Using such auditory hierarchical stimuli, attentional persistence occurs independently of target pattern, response and absolute scope repetition (Justus and List, 2005; List and Justus, 2010). Because attentional persistence to scope manifests for both vision and audition, one question is whether attentional persistence to a hierarchical level can occur across modalities. In other words, might attending to global auditory information bias an individual toward global visual information, and vice versa? Similarly, might attending to local information in one modality prime subsequent local processing in the other? In one study, Bouvet et al. (2011) showed that *unimodal* auditory and visual biases in attentional scope were correlated in individuals. Nevertheless, no study has directly assessed trial-by-trial cross-modal priming, which more directly addresses a potentially shared (or interactive) scope selection mechanism across vision and audition.

There is already evidence that attention operates across the auditory and visual modalities for spatial attention to locations (but not scope; Driver and Spence, 1998a,b). For example, Spence and Driver (1997) showed that exogenous spatial auditory cues influenced visual discrimination (but not vice versa). Their later work extended these findings to also reveal bi-directional auditory-visual inhibition of return effects (Spence and Driver, 1998; Spence et al., 2000). Other research has further delineated certain limits on cross-modal attention effects, especially for endogenous attention (e.g., Ward et al., 2000; Soto-Faraco et al., 2005; Prime et al., 2008; Ahveninen et al., 2019; though see Spence and McDonald, 2004). Studies on brain-injured individuals have demonstrated attentional independence in the auditory and visual systems by, for instance, dissociating auditory and visual spatial attention deficits (extinction and hemi-spatial neglect; e.g., Sinnett et al., 2007; Barrett et al., 2010; cf., Rapp and Hendel, 2003; Brozzoli et al., 2006; Jacobs et al., 2012). Despite demonstrations that auditory-visual spatial attention *can* be dissociated, it yet remains unknown whether the process of attending to local or global information interacts across modalities, as it can, especially, in certain exogenous situations (Spence and Driver, 1998; Spence et al., 2000; Spence and McDonald, 2004). Therefore, although research has not yet determined whether a cross-modal shared or interactive mechanism might exist for attentional scope, it is at least plausible that attentional selection of scope could operate across the visual and auditory modalities in a multi-modal context.

Therefore, in the current experiment, visual and auditory hierarchical stimuli were intermixed to test the hypothesis that the adopted attentional scope (local or global) in one modality (visual or auditory) would prime individuals to persist at that scope (local or global) in the other modality (auditory or visual). Critically, target level (local or global) and modality (auditory or visual) were unpredictable from one trial to the next. Additionally, by using analogous visual and auditory hierarchical stimuli, participants were tasked with identifying a pattern regardless of its level or modality (**Figure 1**). Each hierarchical stimulus was either a nine circle (visual) or tone (auditory) stimulus arranged such that each local pattern (composed of three circles or tones, respectively) was repeated three times and organized to form a global pattern. One additional important benefit of using these stimulus sets is that, unlike commonly-used alphanumeric stimuli (e.g., Navon, 1977), both the local and global patterns require grouping (List et al., 2013). Whether auditory or visual, local and global stimuli were three-element patterns (**Figure 1**). Participants could therefore respond to their two assigned target patterns independent of modality and hierarchical level. Should participants show a level-specific priming effect from vision to audition, or vice versa, independent of any

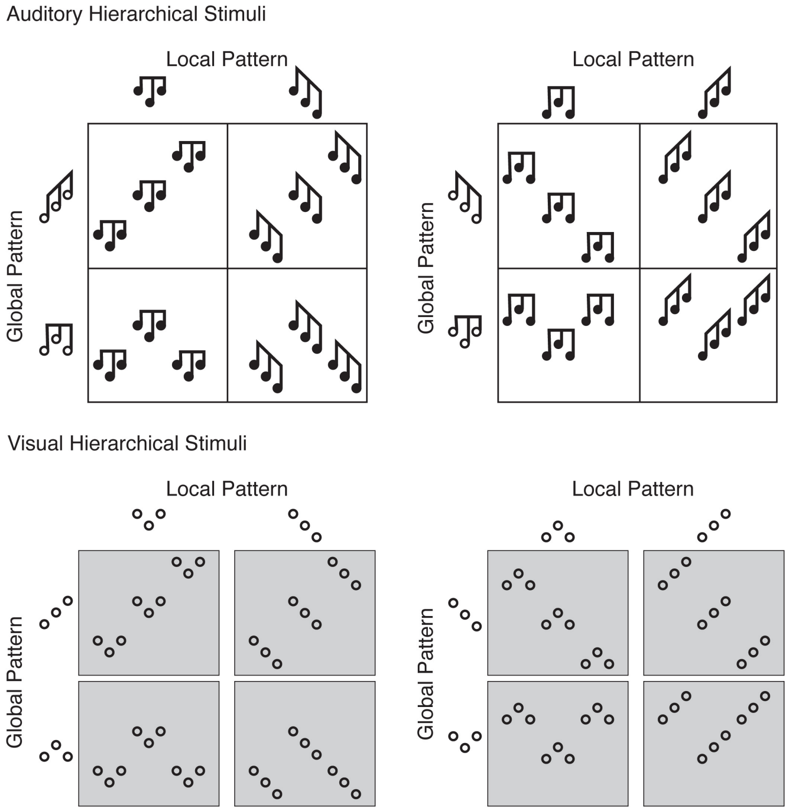


FIGURE 1
The 16 hierarchical stimuli, created by the factorial combination of rising-rising and rising-falling patterns, with falling-rising and falling-falling patterns, and hierarchical level (global and local). For each participant, because their response mapping was counter-balanced, one participant's target pattern (e.g., a rising-rising global pattern is a target for those with rising-rising and rising-falling assignments) was a distracter pattern for another participant (the same rising-rising global pattern is a distracter for those with falling-rising and falling-falling assignments, which would be presented locally). **(Top)** For the auditory hierarchical stimuli, the black notes depict individual tones, where the horizontal axis shows time (the leftmost occurs first and proceeds left to right) and the vertical axis shows frequency (the lowest corresponds to F#$_3$ and highest to A#$_4$). Each local pattern is a three-tone sequence repeated three times to produce a nine-tone global pattern. If a participant were assigned a response mapping of falling-rising and falling-falling patterns, they should respond "falling-rising" when the two leftmost stimuli are presented (falling-rising pattern presented at the local level) as well as when the two bottom right stimuli are presented (falling-rising pattern presented at the global level). **(Bottom)** For the visual hierarchical stimuli, "rising" refers to a southwest-to-northeast left-to-right relationship between neighboring circles, and "falling" refers to a northwest-to-southeast left-to-right relationship between neighboring circles. As with the auditory stimuli, if a participant were assigned a response mapping of falling-rising and falling-falling patterns, they should respond "falling-rising" when the two leftmost stimuli are presented (falling-rising pattern presented at the local level) as well as when the two bottom right stimuli are presented (falling-rising pattern presented at the global level). For those more familiar with visual hierarchical letter stimuli, the analogy is presenting the letters A and E at one level, and S and H at another level. Participants respond to, for example, A and S as target patterns, whether presented at the local or global level.

response or target priming, this would support a shared (or at least interactive) attentional mechanism for selecting auditory and visual scope.

2. Materials and methods

2.1. Participants

As in previous studies using these auditory hierarchical stimuli (e.g., Justus and List, 2005; List and Justus, 2010), right-handed participants reporting 5 or more years of musical experience were recruited. Of the 48 who volunteered, 24 reached the practice criteria described below (13 women; 11 men; M = 19.96 years, SD = 1.40). All participants were undergraduate students who were compensated financially or with course extra credit. All participants provided written informed consent before participating (Hamilton College IRB# SP14-112).

2.2. Stimuli

2.2.1. Auditory stimuli

Auditory hierarchical stimuli were as in Justus and List (2005), Experiment 2. Each 100-ms tone had 10-ms on and off ramps, comprised five 1/n amplitude harmonics, with fundamental frequencies in nine whole-steps ranging from F#$_3$-A#$_4$. Stimuli were presented at ~72 dB SPL through Sennheiser HD280 headphones during the practice and experiment.

Hierarchical stimuli were created by sequencing nine tones without inter-stimulus intervals (**Figure 1**, top). Each local pattern comprised three tones presented in a falling-rising, falling-falling, rising-falling, or rising-rising sequence. Each global pattern comprised three local patterns presented in a falling-rising, falling-falling, rising-falling, or rising-rising sequence. As is shown in the top of **Figure 1**, a factorial combination of (falling-rising, falling-falling) by (rising-falling or rising-rising) by level (global, local) resulted in eight auditory hierarchical stimuli. In this way, participants are always only able to accurately identify one of their two assigned target patterns (either falling-rising and falling-falling, or rising-falling and rising-rising) at the local or global level. The distractor pattern (rising-falling or rising-rising, or falling-rising or falling-falling, respectively) occurs necessarily at the other level. By counter-balancing target patterns across the final group of 24 participants, the same stimulus serves as a local target trial for one group of participants and a global target trial for another group of participants.

2.2.2. Visual stimuli

Visual hierarchical stimuli were as in List et al. (2013); unfilled stimulus set; **Figure 1**, bottom. Black visual hierarchical stimuli were centered on a white background, and comprised nine circle outlines (0.6°-diameter; 0.1° linewidth) spanning a maximum of 7.2° × 7.2° for a whole nine-circle hierarchical figure, with local patterns spanning 1.9° × 1.9° maximum. A black filled circle (0.2°-diameter) served as fixation.

As in the auditory hierarchical stimuli, nine elements were arranged to create the visual hierarchical stimuli (**Figure 1**, bottom). Each local pattern comprised three circles presented in a falling-rising [∨], falling-falling [\], rising-falling [∧], or rising-rising [/] sequence (where the spatial relation between two circles is described as rising, a southwest to northeast direction, and falling, a northwest to southeast direction). Each global pattern comprised three local patterns presented in a falling-rising, falling-falling, rising-falling, or rising-rising sequence. As is shown in **Figure 1** (bottom), a factorial combination of (falling-rising, falling-falling) by (rising-falling or rising-rising) by level (local, global) resulted in eight visual hierarchical stimuli.

2.3. Procedure

Participants were seated ∼57 cm from a 1,920 × 1,080 resolution monitor, running at 60 Hz. Each participant was assigned two target patterns, either rising-rising and rising-falling, or falling-falling and falling-rising (**Figure 1**). Each pattern was depicted as a triplet on response box keys, and participants were instructed to respond as quickly and accurately with their right and left index fingers (pattern-side mappings were counter-balanced across the final group of 24 participants). Participants reported which of their two target patterns was presented, *regardless of the level at which it occurred (global or local), or in which modality it occurred.* This is referred to as a divided-attention task in the literature (e.g., Lamb and Robertson, 1989; Hübner et al., 2007), because participants must identify one of their two target patterns without knowing prior to stimulus presentation whether it will be presented globally or locally, or auditorily or visually.

Because the auditory task is more challenging for participants to learn than the visual task, participants were familiarized with the auditory task first. They were presented with auditory examples of each target triplet alone (e.g., falling-falling and falling-rising) at a fast and a slow rate, and were given unlimited time to respond by pressing the buttons. They were then presented with eight randomly interleaved trials to further practice the task and response mapping.

Participants were then shown a visual depiction of the full array of hierarchical stimuli (akin to **Figure 1**) and were explicitly shown their target pattern in each of the stimuli. In order to continue to the experiment, participants were required to reach a minimum of 14/16 trials correct within six practice blocks. If participants reached criterion performance in the auditory practice, practice with the visual stimuli alone followed, and if they reached the same criterion responding to visual hierarchical stimuli, then the multi-modal auditory-visual practice followed. The 24 participants who reached the criteria to participate in the experiment completed a mean of 4.1 (*SD* = 1.5) auditory, 1.8 (*SD* = 1.3) visual and 2.3 (*SD* = 1.1) multi-modal practice blocks. Due to experimenter error, four participants' unimodal auditory and visual practice data are missing. Participants were encouraged to ask questions between practice blocks, and to focus on responding both as accurately and as quickly as possible.

In the experiment, participants completed six blocks of 65 trials apiece, with self-paced breaks between blocks. In each block, 64 trials were sequenced so that each trial (according to its target pattern, target level and target modality) followed each other trial type once to balance priming repetition and changes. Because there were two target patterns, two levels and two modalities, eight trial types resulted which followed each of the eight trial types once (8^2 = 64 trials). However, because the first trial is not subject to priming itself, it was repeated at the end to include it in the priming analyses (hence, 65 trials per block). Each of the 24 final participants completed six distinct fixed trial orders, and block order was varied using a partial Latin-squares design.

Figure 2 shows an example sequence of three trials. Each auditory trial began with a central fixation dot that was presented for 1.9 s. After 1 s of fixation, a 900-ms hierarchical auditory stimulus was presented. Visual trials also began with 1 s of fixation, followed by a visual hierarchical stimulus for 100 ms, and fixation for 800 ms. Blank and silent 1.5-s inter-trial intervals (ITIs) separated all trials. Responses could occur any time from stimulus presentation until the next trial began (i.e., within 2,400 ms of stimulus onset).

Trials were coded for modality and target level, and to enable analysis of the priming effects, trials were coded according to transitions between N and N-1 target patterns, target levels, distractor pattern and modality. In **Figure 2**, assuming a participant is assigned falling-rising and falling-falling target patterns, they would respond falling-rising on the first trial, falling-rising on the second trial and falling-falling on the third trial. In terms of priming, the second trial is an example of within-modality (visual → *visual*), same-target pattern (i.e., same-response), same-level (local → *local*) and different-distractor pattern (rising-rising → rising-falling). In terms of priming, the third trial is an example of across-modality (visual → *auditory*), different-target pattern (i.e., different-response), different-level (local → *global*) and same-distractor pattern (rising-falling → rising-falling) trial. The design specifies the current trial's modality (visual, auditory), target level (global, local), as well as its relation to the previous trial: modality priming (same, different), target priming (same, different) and level priming (same, different).

2.4. Data analysis

To demonstrate level-specific attentional persistence independent of target and response priming, it is critical to compare certain conditions *a priori* (as in, e.g., List and Justus, 2010; List et al., 2013). Namely, level-specific priming is demonstrated by showing that same-level responses are facilitated relative to different-level responses, *when the target and response change*. Otherwise, the priming effect would be conflated with target (and/or response) priming. For example, to claim true auditory level-specific priming

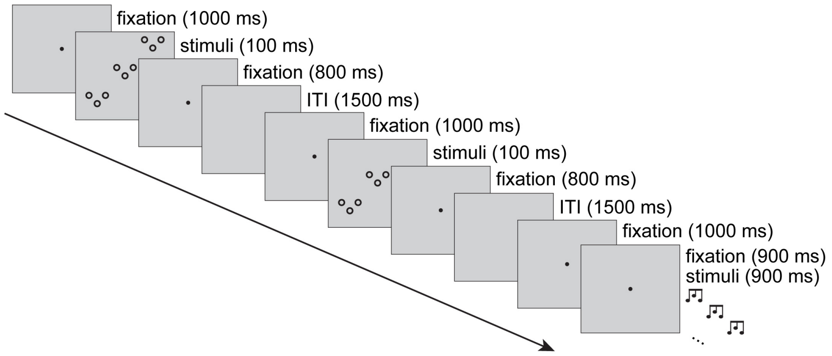


FIGURE 2

The experimental procedure displayed with a sequence of three trials. Visual and auditory trials were intermixed, with 1.5-s inter-trial intervals (ITIs). In the figure, the first trial shows a visual rising-rising global pattern and falling-rising local pattern. The second trial shows a visual rising-falling global pattern and a falling-rising local pattern. The third trial shows an auditory falling-falling global pattern and a rising-falling local pattern (while a simultaneous visual fixation is on the screen). Trials were coded according to modality, modality priming, target priming, level, and level priming. Thus, for a participant assigned falling-rising and falling-falling target patterns, in the examples depicted, accurate responses would be falling-rising, falling-rising and falling-falling. The second trial (subject to priming from the first) would contribute to the visual, within-modality, same-target, local, same-level condition, and the third trial (subject to priming from the second) would contribute to the auditory, across-modality, different-target, global, different-level condition.

of vision, a reliable difference would need to manifest between the same- and different-level responses in the across-modality, visual, and different-target condition. Identity priming, on the other hand, is measured by comparing performance on same- and different-level trials for the *same* target pattern. For identity priming, the target pattern (and response) are held constant, and the comparison is between repeated and changed hierarchical level. Therefore, eight planned paired-samples *t*-tests were conducted for same vs. different level, for auditory trials, within or across modality, and for visual trials, within or across modality (**Figure 3**). Effect sizes (as Cohen's *d)* and Bayes factors (*K)* are also reported for these analyses. The omnibus analysis and follow up analyses are depicted in **Figures 4–7**, and the ANOVA table is provided in the **Supplementary material**.

3. Results

Accurate trials' ($M = 88.4\%$, $SD = 4.4$; excludes both misses and errors, as well as trials following misses or errors for priming analyses) response times (RTs) were trimmed, removing outliers ±3 *SD*s, and submitted to planned paired comparisons (see section "2.4 Data analysis" above). For completeness, an omnibus repeated-measures ANOVA was also conducted with modality (auditory, visual) × modality priming (same, different) × level (local, global) × level priming (same, different) × target pattern priming (same, different) as factors. The priming factors reflect the coding of trial N, relative to trial N-1 (**Figure 2**).

3.1. Cross-modal level-priming

The primary findings are illustrated in **Figure 3**: no level-specific priming occurred, within or across modalities. In different visual target trials, no same- vs. different-level benefit was found within-modality or across modalities, $ts < 1$, $Ks > 3$ (moderate evidence for null), nor was it present for auditory within-modality trials, $t < 1$, $Ks > 3$ (moderate evidence for null), and an opposite (same > different-level) effect emerged for auditory across-modality, 37 ms, $t(23) = 2.10$, $p = 0.047$, Cohen's $d = 0.43$, $K = 0.09$ (anecdotal evidence for difference). However, this latter result did not survive Bonferroni ($p < 0.00625$) or Bonferroni-Holm correction for multiple comparisons, whereas the following three effects did. Despite a lack of level-specific priming, identity priming emerged in three cases: Visual within-modality, 62 ms, $t(23) = 3.907$, $p = 0.001$, Cohen's $d = 0.81$, $K = 0.02$ (very strong evidence for difference); visual across-modality, 37 ms, $t(23) = 4.107$, $p < 0.001$, Cohen's $d = 0.84$, $K = 0.01$ (very strong evidence for difference); and auditory within-modality, $t(23) = 4.365$, $p < 0.001$, Cohen's $d = 0.89$, $K = 0.008$ (extreme evidence for difference). These robust identity priming effects provide confidence that participants were engaged in the task, show that the experimental design was rigorous enough to detect priming effects, and rule out potential RT ceiling and floor limits in detecting level-priming effects. For the auditory across-modality trials, no reliable identity priming emerged, $t(23) = 1.344$, $p = 0.192$, Cohen's $d = 0.28$, $K = 2.74$ (anecdotal evidence for null).

3.2. Omnibus ANOVA

In the omnibus ANOVA (see **Supplementary Table 1** for full results), four main effects emerged. Participants responded 512 ms slower to auditory than visual targets, $F(1, 23) = 50.372$, $p < 0.001$, $\eta_p^2 = 0.94$. Participants were also 46 ms slower to respond when modality switched rather than repeated, $F(1, 23) = 39.709$, $p < 0.001$, $\eta_p^2 = 0.63$. Overall, RTs were 36 ms faster to global than local targets, $F(1, 23) = 4.329$, $p = 0.049$, $\eta_p^2 = 0.16$. Lastly, participants' RTs were 25 ms faster for same-level compared to changed-level, $F(1, 23) = 17.391$, $p < 0.001$, $\eta_p^2 = 0.43$. No main effect was found for target priming, $F(1, 23) = 2.408$, $p = 0.13$, $\eta_p^2 = 0.10$. All the reliable main effects, barring Level, remained even when adopting a strict Bonferroni or Bonferroni-Holm correction for multiple comparisons.

The omnibus ANOVA also revealed six higher-order interactions that are described below, as well as follow-up paired *t*-tests. In support of the reported planned comparisons above, an overall target

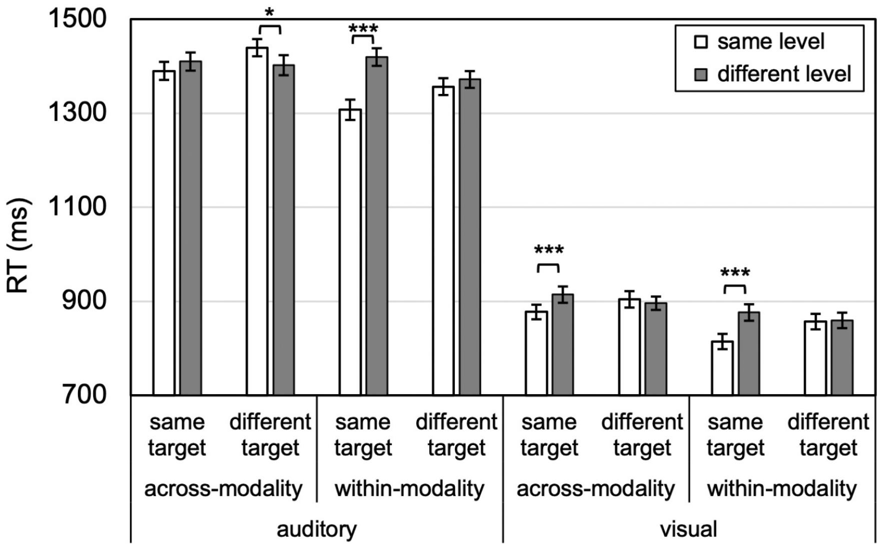


FIGURE 3
Attentional persistence to level is absent (for different targets, no benefit for same-level compared to different-level), whereas identity priming is more prevalent (for same targets, compare same- to different-level). Error bars reflect *SE*s adjusted for within-subjects comparisons, $^{*}p < 0.05$, $^{***}p \leq 0.001$.

priming by level priming interaction emerged, $F(1, 23) = 23.324$, $p < 0.001$, $\eta_p^2 = 0.50$ (**Figure 4**). By comparing same- vs. different-level RTs, no level-priming occurred when the target changed, 7 ms (different faster than same), $t < 1$, Cohen's $d = 0.19$, whereas identity priming did occur when the target repeated, 58 ms, $t(23) = 5.593$, $p < 0.001$, Cohen's $d = 1.14$ (**Figure 4**).

A two-way interaction between modality priming and level priming, $F(1, 23) = 24.485$, $p < 0.001$, $\eta_p^2 = 0.52$, was qualified by three-way interaction between modality, modality priming and level priming, $F(1, 23) = 4.496$, $p = 0.045$, $\eta_p^2 = 0.16$ (**Figure 5**). When modality repeated, both auditory and visual targets were faster for same- than different-level, auditory: 64 ms, $t(23) = 4.107$, $p < 0.001$, Cohen's $d = 0.84$; visual: 32 ms, $t(23) = 2.807$, $p = 0.01$, Cohen's $d = 0.57$. When modality changed, however, visual responses showed a trend for benefit for same vs. different-level, 14 ms, $t(23) = 1.991$, $p = 0.06$:, Cohen's $d = 0.41$, whereas auditory did not, 9 ms (different faster than same), $t < 1$, Cohen's $d = 0.17$. Thus, a benefit for level-repetition (with identity and level-priming conflated) was only evident for within-modality transitions.

A two-way interaction between modality and level, $F(1, 23) = 17.249$, $p < 0.001$, $\eta_p^2 = 0.43$, was qualified by three-way interaction between modality, modality priming and level, $F(1, 23) = 6.156$, $p = 0.021$, $\eta_p^2 = 0.21$ (**Figure 6**). A 76-ms global advantage was present in auditory trials, $t(23) = 3.121$, $p = 0.005$, Cohen's $d = 0.64$, but not visual trials, 5 ms, $t < 1$. The three way interaction was due to this auditory global advantage being greater for within-modality, 93 ms, $t(23) = 3.777$, $p = 0.001$, Cohen's $d = 0.77$, compared with across-modality transitions, 14 ms, $t(23) = 2.133$, $p = 0.04$, Cohen's $d = 0.44$. In brief, an auditory global (over local) advantage was present overall, though most evident for within-modality transitions.

Lastly, a three-way interaction between level, modality priming and target priming emerged, $F(1, 23) = 6.077$, $p = 0.022$, $\eta_p^2 = 0.21$ (**Figure 7**). Across-modality, a global advantage was absent for repeated targets, 28 ms, $t(23) = 1.604$, $p = 0.12$, Cohen's $d = 0.33$, but present for changed targets, 52 ms, $t(23) = 2.669$, $p = 0.01$, Cohen's $d = 0.55$. The reverse was true for within-modality transitions: a global advantage was present for repeated targets, 44 ms, $t(23) = 2.307$, $p = 0.03$, Cohen's $d = 0.47$, but absent (19 ms) for changed targets, $t < 1$, Cohen's $d = 0.18$. No other higher order interactions reached significance.

4. Discussion

The primary result from this study is that no cross-modal attentional persistence for scope occurred—participants did not benefit from targets being presented locally (or globally) for subsequent local (or global) targets when switching from vision

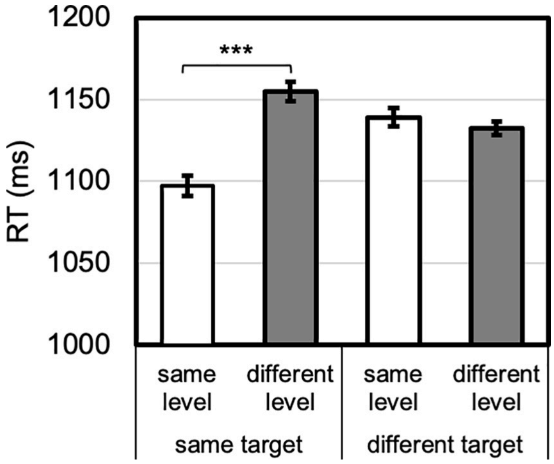


FIGURE 4
Level priming by target priming interaction. Identity priming was present (same vs. different level for same target), whereas level-priming was not (same vs. different level for different target). Error bars reflect *SE*s adjusted for within-subjects comparisons, $^{***}p < 0.001$.

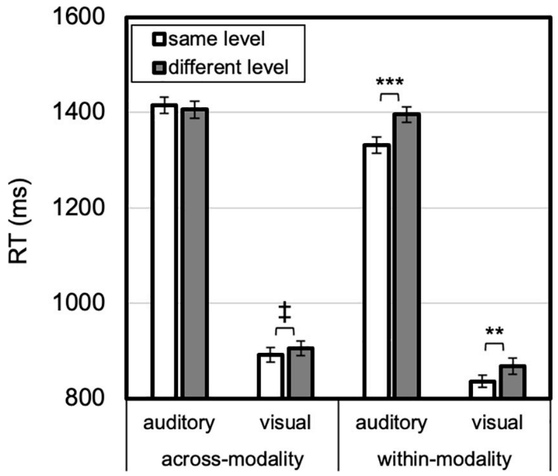


FIGURE 5
Modality by modality priming by level priming. A within-modality same-level benefit (vs. different-level) was greater for auditory than visual targets. A trend for a visual across-modality same-level benefit was present. Error bars reflect *SE*s adjusted for within-subjects comparisons, $^{\ddagger}p < 0.10$, $^{**}p \le 0.01$, $^{***}p \le 0.001$.

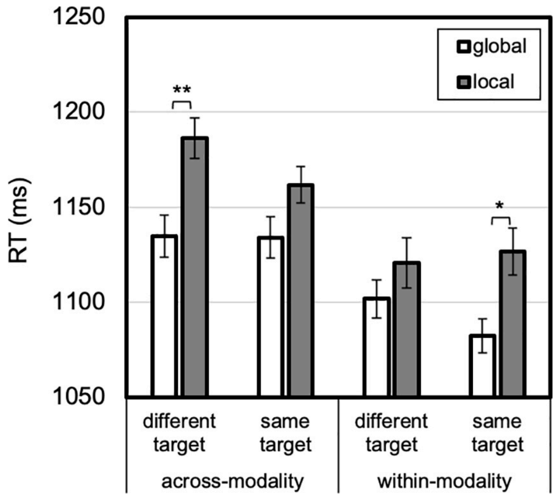


FIGURE 7
Level by modality priming by target priming. A global advantage was present for different targets after modality changed and for repeated targets after modality repeated. Error bars reflect *SE*s adjusted for within-subjects comparisons, $^{*}p < 0.05$, $^{**}p \le 0.001$.

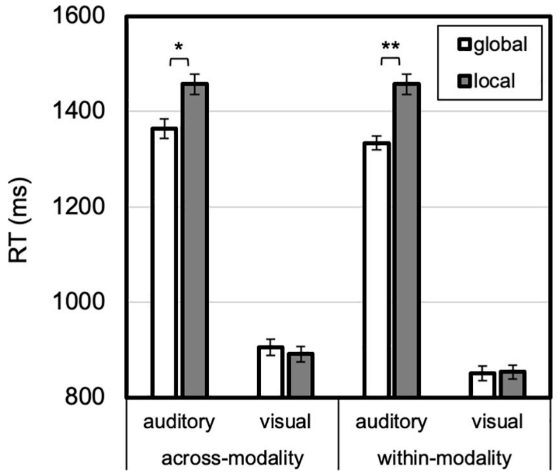


FIGURE 6
Level by modality priming by modality. An auditory global (vs. local) advantage was larger after modality repeated than when modality changed. Error bars reflect *SE*s adjusted for within-subjects comparisons, $^{*}p < 0.05$, $^{**}p \le 0.001$.

to audition, or vice versa (**Figure 3**). Surprisingly, and contrary to previous *unimodal* auditory and visual studies (e.g., Ward, 1982; Robertson, 1996; Filoteo et al., 2001; Justus and List, 2005; List and Justus, 2010; List et al., 2013), the multi-modal context also eliminated within-modality level-specific priming, as supported by the Bayes factor showing moderate evidence for the null hypothesis. Even when visual (or auditory) targets followed other visual (or auditory) targets, the typical benefit for repeating a target's local or global level was absent. Thus, the multi-modal context interfered with unimodal attentional settings that typically drive persistence in processing local or global information. Paradoxically, the disruption of *unimodal* scope priming suggests that the multi-modal context has an impact on level-specific attentional persistence. These data are therefore inconsistent with fully independent visual and auditory attentional systems—were the systems entirely independent, cross-modal level-specific priming would not emerge, but within-modality level-specific priming should still manifest. The data are also inconsistent with the hypothesized level-specific cross-modal interactions, because none emerged. Instead, the data point to a goal-directed or strategic cross-modal interaction whereby maintaining attentional flexibility across modalities with distinct hierarchical levels has as its consequence the elimination of unimodal level-specific priming.

Critically, however, not all priming effects were eliminated. Identity priming (reflected as an advantage for repeated level vs. changed level in repeated target/response trials) was present in three cases, and the Bayes factor revealed very strong to extreme support for a difference between conditions. Namely, for visual trials, whether preceded by auditory or visual trials, participants showed a benefit for the target pattern to repeat at the same level rather than change levels. This was also true for within-modality auditory trials. These results are important because they establish the rigor of the method in detecting priming effects, whether for visual or auditory targets (whose RTs do differ considerably). These identity priming results suggest that the null level-priming effects are not simply due to, for example, poor execution or unmotivated participants—otherwise, neither would be present.

Because the absence of unimodal level-priming effects was unexpected, it is important to consider how the multi-modal context may have disrupted attentional persistence across modalities. One consideration is whether presenting stimuli in both visual and auditory modalities created an additional load on participants compared with prior unimodal studies. Indeed, participants were required to process more and different kinds of stimuli. However, a few points challenge a (simple) load argument. First, all participants were required to practice until meeting a uniform minimum level of accuracy within each modality and in a multi-modal context. Therefore, commensurate with previous unimodal auditory studies in which level-priming occurred (e.g., Justus and List, 2005; List and Justus, 2010), a baseline level of accuracy was achieved. Second, when comparing the accuracy rates and RTs from this study to other unimodal studies showing level-priming effects (e.g., Robertson, 1996; Kim et al., 1999; Justus and List, 2005; List and Justus, 2010; List et al., 2013), performance is well-matched for each modality. Third, the multi-modal context only affected level-priming, and not identity priming. Any argument that load-related difficulty eliminated priming effects would need to account for why identity priming would be spared, whereas level-priming effects would be *selectively* eliminated. Although the current study cannot rule out the possibility that other load manipulations might have similar selective consequences, it is at least established that when participants are required to allocate attention flexibly across visual and auditory scope, there is no evidence that they derive a benefit from repetition of attentional selection within or across modalities.

So why were cross-modal level-specific interactions not found? One possibility is that they will never occur. However, from previous studies (see, e.g., the debate between Spence and Driver, and Ward and his colleagues referenced in the Introduction), variations in stimulus and task parameters can substantively affect whether cross-modal effects are observed. It may yet be possible that level-specific attentional persistence across modalities might occur with variations in methodological approach. One candidate stimulus change is drawn from the work of Ivry and Robertson (1998) and Robertson and Ivry (2000). They surveyed a broad range of research on hierarchical processing, and proposed an information processing theory, the double filtering by frequency theory. The theory holds that an initial attentional selection of relevant frequency information occurs in both vision and audition, and that a subsequent second stage involves the attentional filtering of relatively higher and lower frequencies in left and right hemispheres, respectively. There is ample evidence that visual spatial frequency selection is what underlies, or at least depends on similar mechanisms as, attentional selection of local or global information (e.g., Shulman et al., 1986; Shulman and Wilson, 1987; Robertson, 1996; Flevaris et al., 2011). Furthermore, processing of auditory frequency information has been shown to reflect similar hemispheric asymmetries (Ivry and Lebby, 1993) to those engaged in processing visual spatial frequencies (e.g., Kitterle et al., 1990). Thus, to observe cross-modal level-specific priming, it may be important that the auditory and visual stimuli be better matched by both requiring frequency selection (e.g., using the stimuli from Justus and List, 2005, Experiment 1). In the current study, though local and global auditory selection *could* be based on frequency information, the patterns vary over time as well, and therefore participants could use both the frequency and temporal dimensions to make their decisions. Thus, in this study, it may be that this additional auditory temporal dimension interfered with cross-modal interactions that might otherwise occur when only frequency-based selection is possible. One compelling piece of evidence supports the importance of frequency selection in producing level-specific priming: Robertson (1996) manipulated the spatial frequency content in visual hierarchical stimuli, and showed that level-based priming effects were eliminated (whereas, importantly, they occurred under other stimulus manipulations).

Another possible avenue for future research into cross-modal hierarchical processing is to match auditory and visual stimuli on the basis of a spatial hierarchy. Some evidence points to the necessity that space be relevant in both modalities for cross-modal effects to emerge (e.g., Spence and McDonald, 2004). In the current study, space was only relevant in the visual modality, because by adopting Justus and List's (2005, Experiment 2) stimulus set, this study inherently adopts their arguments that frequency and time are the relevant auditory dimensions for local and global selection (also relying indirectly on Kubovy and Van Valkenburg's (2001) auditory object attributes). Although spatial locations may seem evidently analogous across vision and audition, in multi-modal research, a ubiquitous and persisting problem is understanding which dimensions might be analogous across modalities and how flexible these mappings are (e.g., Marks, 1974; Evans and Treisman, 2010). For instance, even though space is common to multiple modalities, visual space can map to multiple auditory dimensions (e.g., auditory space or frequency). Nevertheless, because auditory hierarchical stimuli varying in frequency and time did not interact with visual spatial hierarchical stimuli here, it would be worth further investigating cross-modal interactions with both auditory and visual *spatial* hierarchical stimuli—under those conditions, stimuli in both modalities would be able to be parsed spatially into local and global levels, potentially providing an even stronger opportunity for cross-modal interactions to arise.

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 Hamilton College Institutional Review Board (SP14-112). The patients/participants provided their written informed consent to participate in this study.

Author contributions

AL was responsible for the reported research and writing.

Funding

Hamilton College supported the reported research.

Acknowledgments

Special thanks to Sandhya Rao, Elizabeth Prescott, Katherine Callahan, and Emma Feitelson for help in conducting portions of this study. Thanks to Timothy Justus for collaborations leading up to this research.

Conflict of interest

The author declares 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/fnhum.2022.1043475/full#supplementary-material

References

Ahveninen, J., Ingalls, G., Yildirim, F., Calabro, F. J., and Vaina, L. M. (2019). Peripheral visual localization is degraded by globally incongruent auditory-spatial attention cues. *Exp. Brain Res.* 237, 2137–2143. doi: 10.1007/s00221-019-05578-z

Barrett, D. J., Edmondson-Jones, A. M., and Hall, D. A. (2010). Attention in neglect and extinction: Assessing the degree of correspondence between visual and auditory impairments using matched tasks. *J. Clin. Exp. Neuropsychol.* 32, 71–80. doi: 10.1080/13803390902838058

Bouvet, L., Rousset, S., Valdois, S., and Donnadieu, S. (2011). Global precedence effect in audition and vision: Evidence for similar cognitive styles across modalities. *Acta Psychol.* 138, 329–335. doi: 10.1016/j.actpsy.2011.08.004

Brozzoli, C., Demattè, M. L., Pavani, F., Frassinetti, F., and Farnè, A. (2006). Neglect and extinction: Within and between sensory modalities. *Restor. Neurol. Neurosci.* 24, 217–232.

Bultitude, J. H., and Rafal, R. D. (2010). Amelioration of right spatial neglect after visuo-motor adaptation to leftward-shifting prisms. *Cortex* 46, 404–406. doi: 10.1016/j.cortex.2009.06.002

Bultitude, J. H., Rafal, R. D., and List, A. (2009). Prism adaptation reverses the local processing bias in patients with right temporo-parietal junction lesions. *Brain* 132, 1669–1677. doi: 10.1093/brain/awp096

Delis, D. C., Robertson, L. C., and Efron, R. (1986). Hemispheric specialization of memory for visual hierarchical stimuli. *Neuropsychologia* 24, 205–214. doi: 10.1016/0028-3932(86)90053-9

Driver, J., and Spence, C. (1998a). Attention and the crossmodal construction of space. *Trends Cogn. Sci.* 2, 254–262. doi: 10.1016/S1364-6613(98)01188-7

Driver, J., and Spence, C. (1998b). Cross-modal links in spatial attention. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 353, 1319–1331. doi: 10.1098/rstb.1998.0286

Evans, K. K., and Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *J. Vis.* 10:6. doi: 10.1167/10.1.6

Filoteo, J. V., Friedrich, F. J., and Stricker, J. L. (2001). Shifting attention to different levels within global-local stimuli: A study of normal participants and a patient with temporal-parietal lobe damage. *Cogn. Neuropsychol.* 18, 227–261. doi: 10.1080/02643290125839

Flevaris, A. V., Bentin, S., and Robertson, L. C. (2011). Attention to hierarchical level influences attentional selection of spatial scale. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 12–22. doi: 10.1037/a0019251

Hübner, R., Volberg, G., and Studer, T. (2007). Hemispheric differences for global/local processing in divided attention tasks: Further evidence for the integration theory. *Percept. Psychophys.* 69, 413–421. doi: 10.3758/BF03193762

Ivry, R., and Lebby, P. (1993). Hemispheric differences in auditory perception are similar to those found in visual perception. *Psychol. Sci.* 4, 41–45. doi: 10.1111/j.1467-9280.1993.tb00554.x

Ivry, R., and Robertson, L. (1998). *The two sides of perception*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/6942.001.0001

Jacobs, S., Brozzoli, C., and Farnè, A. (2012). Neglect: A multisensory deficit? *Neuropsychologia* 50, 1029–1044. doi: 10.1016/j.neuropsychologia.2012.03.018

Justus, T., and List, A. (2005). Auditory attention to frequency and time: An analogy to visual local-global stimuli. *Cognition* 98, 31–51. doi: 10.1016/j.cognition.2004.11.001

Kim, N., Ivry, R. B., and Robertson, L. C. (1999). Sequential priming in hierarchically organized figures: Effects of target level and target resolution. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 715–729. doi: 10.1037/0096-1523.25.3.715

Kimchi, R., and Palmer, S. E. (1982). Form and texture in hierarchically constructed patterns. *J. Exp. Psychol. Hum. Percept. Perform.* 8, 521–535. doi: 10.1037/0096-1523.8.4.521

Kinchla, R. A., and Wolfe, J. M. (1979). The order of visual processing: "Top-down," "bottom-up", or "middle-out". *Percept. Psychophys.* 25, 225–231. doi: 10.3758/BF03202991

Kitterle, F. L., Christman, S., and Hellige, J. B. (1990). Hemispheric differences are found in the identification, but not the detection, of low versus high spatial frequencies. *Percept. Psychophys.* 48, 297–306. doi: 10.3758/BF03206680

Kubovy, M., and Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition* 80, 97–126. doi: 10.1016/S0010-0277(00)00155-4

Lamb, M. R., and Robertson, L. C. (1988). The processing of hierarchical stimuli: Effects of retinal locus, locational uncertainty, and stimulus identity. *Percept. Psychophys.* 44, 172–181. doi: 10.3758/BF03208710

Lamb, M. R., and Robertson, L. C. (1989). Do response time advantage and interference reflect the order of processing of global- and local-level information? *Percept. Psychophys.* 46, 254–258. doi: 10.3758/BF03208087

Lamb, M. R., and Robertson, L. C. (1990). The effect of visual angle on global and local reaction times depends on the set of visual angles presented. *Percept. Psychophys.* 47, 489–496. doi: 10.3758/BF03208182

Lamb, M. R., Robertson, L. C., and Knight, R. T. (1988). Effects of lesions of temporal-parietal junction on perceptual and attentional processing in humans. *J. Neurosci.* 8, 3757–3769.

Lamb, M. R., Robertson, L. C., and Knight, R. T. (1989). Attention and interference in the processing of global and local information: Effects of unilateral temporal-parietal junction lesions. *Neuropsychologia* 27, 471–483. doi: 10.1016/0028-3932(89)90052-3

List, A., and Justus, T. (2010). Relative priming of temporal local–global levels in auditory hierarchical stimuli. *Atten. Percept. Psychophys.* 72, 193–208. doi: 10.3758/APP.72.1.193

List, A., Grabowecky, M., and Suzuki, S. (2013). Local and global level-priming occurs for hierarchical stimuli composed of outlined, but not filled-in, elements. *J. Vis.* 13:23. doi: 10.1167/13.2.23

Marks, L. E. (1974). On associations of light and sound: The mediation of brightness, pitch, and loudness. *Am. J. Psychol.* 87, 173–188. doi: 10.2307/1422011

Martin, M. (1979). Local and global processing: The role of sparsity. *Mem. Cogn.* 7, 476–484. doi: 10.3758/BF03198264

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cogn. Psychol.* 9, 353–383. doi: 10.1016/0010-0285(77)90012-3

Ouimet, T., Foster, N. E., and Hyde, K. L. (2012). Auditory global-local processing: Effects of attention and musical experience. *J. Acoust. Soc. Am.* 132, 2536–2544. doi: 10.1121/1.4747009

Prime, D., McDonald, J., Green, J., Ward, L., and Grondin, S. (2008). When cross-modal spatial attention fails. *Can. J. Exp. Psychol.* 62, 192–197. doi: 10.1037/1196-1961.62.3.192

Rafal, R., and Robertson, L. (1995). "The neurology of visual attention," in *The cognitive neurosciences*, ed. M. S. Gazzaniga (Cambridge, MA: The MIT Press), 625–648.

Rapp, B., and Hendel, S. K. (2003). Principles of cross-modal competition: Evidence from deficits of attention. *Psychon. Bull. Rev.* 10, 210–219. doi: 10.3758/BF03196487

Robertson, L. C. (1996). Attentional persistence for features of hierarchical patterns. *J. Exp. Psychol. Gen.* 125, 227–249. doi: 10.1037/0096-3445.125.3.227

Robertson, L. C., Lamb, M. R., and Knight, R. T. (1988). Effects of lesions of temporal-parietal junction on perceptual and attentional processing in humans. *J. Neurosci.* 8, 3757–3769.

Robertson, L., and Ivry, R. (2000). Hemispheric asymmetries: Attention to visual and auditory primitives. *Curr. Dir. Psychol. Sci.* 9, 59–63. doi: 10.1111/1467-8721.00061

Rossetti, Y., Rode, G., Pisella, L., Farné, A., Li, L., Boisson, D., et al. (1998). Prism adaptation to a rightward optical deviation rehabilitates left hemispatial neglect. *Nature* 395, 166–169. doi: 10.1038/25988

Sanders, L. D., and Poeppel, D. (2007). Local and global auditory processing: Behavioral and ERP evidence. *Neuropsychologia* 45, 1172–1186.

Shulman, G. L., Sullivan, M. A., Gish, K., and Sakoda, W. J. (1986). The role of spatial-frequency channels in the perception of local and global structure. *Perception* 15, 259–273. doi: 10.1068/p150259

Shulman, G. L., and Wilson, J. (1987). Spatial frequency and selective attention to local and global information. *Perception* 16, 89–101. doi: 10.1068/p160089

Sinnett, S., Juncadella, M., Rafal, R., Azañón, E., and Soto-Faraco, S. (2007). A dissociation between visual and auditory hemi-inattention: Evidence from temporal order judgements. *Neuropsychologia* 45, 552–560. doi: 10.1016/j.neuropsychologia.2006.03.006

Soto-Faraco, S., Morein-Zamir, S., and Kingstone, A. (2005). On audiovisual spatial synergy: The fragility of the phenomenon. *Percept. Psychophys.* 67, 444–457. doi: 10.3758/BF03193323

Spence, C., and Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Percept. Psychophys.* 59, 1–22. doi: 10.3758/BF03206843

Spence, C., and Driver, J. (1998). Auditory and audiovisual inhibition of return. *Percept. Psychophys.* 60, 125–139. doi: 10.3758/BF03211923

Spence, C., and McDonald, J. (2004). "The crossmodal consequences of the exogenous spatial orienting of attention," in *The handbook of multisensory processes*, eds G. Calvert, C. Spence, and B. E. Stein (Cambridge, MA: MIT Press), 3–25. doi: 10.7551/mitpress/3422.003.0004

Spence, C., Lloyd, D., McGlone, F., Nicholls, M. E., and Driver, J. (2000). Inhibition of return is supramodal: A demonstration between all possible pairings of vision, touch, and audition. *Exp. Brain Res.* 134, 42–48. doi: 10.1007/s002210000442

Ward, L. M. (1982). Determinants of attention to local and global features of visual forms. *J. Exp. Psychol. Hum. Percept. Perform.* 8, 562–581. doi: 10.1037/0096-1523.8.4.562

Ward, L. M., McDonald, J. J., and Lin, D. (2000). On asymmetries in cross-modal spatial attention orienting. *Percept. Psychophys.* 62, 1258–1264. doi: 10.3758/BF03212127

frontiers | Frontiers in Human Neuroscience

TYPE Original Research
PUBLISHED 02 March 2023
DOI 10.3389/fnhum.2023.1058617

OPEN ACCESS

EDITED BY
Alessia Tonelli,
Italian Institute of Technology (IIT), Italy

REVIEWED BY
Michael Beyeler,
University of California, Santa Barbara,
United States
Meike Scheller,
Durham University, United Kingdom

*CORRESPONDENCE
Shira Shvadron
✉ shira.shvadron@post.runi.ac.il

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

SPECIALTY SECTION
This article was submitted to Sensory Neuroscience, a section of the journal Frontiers in Human Neuroscience

RECEIVED 30 September 2022
ACCEPTED 09 January 2023
PUBLISHED 02 March 2023

CITATION
Shvadron S, Snir A, Maimon A, Yizhar O, Harel S, Poradosu K and Amedi A (2023) Shape detection beyond the visual field using a visual-to-auditory sensory augmentation device.
Front. Hum. Neurosci. 17:1058617.
doi: 10.3389/fnhum.2023.1058617

Shape detection beyond the visual field using a visual-to-auditory sensory augmentation device

Shira Shvadron[1,2*†], Adi Snir[1,2†], Amber Maimon[1,2†], Or Yizhar[1,2,3,4], Sapir Harel[1,2], Keinan Poradosu[1,2,5] and Amir Amedi[1,2]

[1]Baruch Ivcher School of Psychology, The Baruch Ivcher Institute for Brain, Cognition, and Technology, Reichman University, Herzliya, Israel, [2]The Ruth and Meir Rosenthal, Brain Imaging Center, Reichman University, Herzliya, Israel, [3]Research Group Adaptive Memory and Decision Making, Max Planck Institute for Human Development, Berlin, Germany, [4]Max Planck Dahlem Campus of Cognition (MPDCC), Max Planck Institute for Human Development, Berlin, Germany, [5]Weizmann Institute of Science, Rehovot, Israel

Current advancements in both technology and science allow us to manipulate our sensory modalities in new and unexpected ways. In the present study, we explore the potential of expanding what we perceive through our natural senses by utilizing a visual-to-auditory sensory substitution device (SSD), the EyeMusic, an algorithm that converts images to sound. The EyeMusic was initially developed to allow blind individuals to create a spatial representation of information arriving from a video feed at a slow sampling rate. In this study, we aimed to use the EyeMusic for the blind areas of sighted individuals. We use it in this initial proof-of-concept study to test the ability of sighted subjects to combine visual information with surrounding auditory sonification representing visual information. Participants in this study were tasked with recognizing and adequately placing the stimuli, using sound to represent the areas outside the standard human visual field. As such, the participants were asked to report shapes' identities as well as their spatial orientation (front/right/back/left), requiring combined visual (90° frontal) and auditory input (the remaining 270°) for the successful performance of the task (content in both vision and audition was presented in a sweeping clockwise motion around the participant). We found that participants were successful at a highly above chance level after a brief 1-h-long session of online training and one on-site training session of an average of 20 min. They could even draw a 2D representation of this image in some cases. Participants could also generalize, recognizing new shapes they were not explicitly trained on. Our findings provide an initial proof of concept indicating that sensory augmentation devices and techniques can potentially be used in combination with natural sensory information in order to expand the natural fields of sensory perception.

KEYWORDS

spatial perception, visual-auditory, sensory substitution, sensory substitution device (SSD), visual-spatial perception, auditory spatial perception, multisensory spatial perception, multisensory perception

Introduction

In humans, vision is unequivocally considered the dominant sense (Colavita, 1974; Hutmacher, 2019). In addition, cumulative evidence has demonstrated that vision tends to dominate the perception of spatial location when presented alongside conflicting information from other senses. This phenomenon is demonstrated by the well-known ventriloquist effect

(Bruns, 2019). A further connection between the senses of vision and audition is demonstrated in the McGurk effect, in which changing visual stimuli impact the auditory stimulus understood to be heard (McGurk and MacDonald, 1976). And yet, the human visual field has a limitation in that it spans 210° (Traquair, 1938; Strasburger, 2020), leaving humans with visual-perceptual blind spots. In addition, a large part of the 210° field of view is peripheral vision (Millodot, 2014), which undergoes a dramatic spatial/temporal discontinuity (Shapiro et al., 2010). On the other hand, the human auditory spatial field encompasses the entire 360° range. As such, our perception of space largely depends on the integration of information from these two crucial senses.

Irrespective of our perception of them, material objects in space are located around us and known to "have an intimate relationship with space" (Casati and Varzi, 1996, p.205). We are constantly tasked with reliably identifying the location at which these objects around us are found. This is where the integration of the senses and multisensory interactions are thought to come into play. It is known that multisensory integration is an acquired process (Gori et al., 2008) and that adults continually update their perceptual systems, calibrating them to their sensory circumstances (Ernst, 2008). Previous studies have taken different approaches as to how observers can recognize visual shapes from auditory cues (Bach-y-Rita et al., 1969; Bach-y-Rita, 1983, 2004; Ptito et al., 2005; Amedi et al., 2007; Striem-Amit et al., 2012; Maidenbaum et al., 2014). There is still an ongoing debate about how vision and audition are integrated for stimuli learned in the adult brain (e.g., Hertz and Amedi, 2015). Prior research has even indicated that cross-modal attenuation (deactivation) can reverse in sensory cortices after training on sensory substitution algorithms, and associative areas can change their sensory response profiles (Hertz and Amedi, 2015). Research indicates multisensory interactions are found in many cortical and subcortical locations (Alais et al., 2010). This considered, the goals of the present study are first and foremost pragmatic, exploring whether and to what extent sighted people can integrate auditory and visual stimuli presented in 360° into a coherent percept.

We perceive the space around us and understand it through shapes and objects. In this respect, shape recognition has been studied widely with visual objects (Milner, 1974; Pietrini et al., 2004; Peelen and Kastner, 2014; Erdogan and Jacobs, 2017). However, when addressing the role of audition in shape perception complementary to or substituting for vision, it has been shown that both sighted and blind observers can process the spatial properties of objects or shapes (Carello et al., 1998; Collignon et al., 2009; Bizley and Cohen, 2013). For example, audition alone can provide information regarding shape curvature (Boyer et al., 2015). In this study, we examine the abilities of sighted people to recognize visual shapes from hearing in a 360° space around their heads, unlike in the studies above, in which shapes are perceived only in the frontal visual field.

The current study utilized the EyeMusic algorithm, a sensory substitution technique that uses a left-to-right sweep-line technique that processes the visual image column by column (Abboud et al., 2014), in combination with spatial audio (Ambisonics), to create a 360° perceivable version of the algorithm named "Vision360." The resulting auditory-rendered musical fragments preserve the image's shape and spatial positioning. In this study, we tested the possibility of combining spatially oriented sensory information to form single or multiple shape percepts while receiving information beginning in the visual field and ending in the auditory field.

We employed the aforementioned procedures to test whether participants without sensory limitations can identify a visual shape or sets of shapes presented to them in a 360° azimuthal orientation around them, thereby building upon auditory perception for enhancing their natural visual field. Moreover, we asked whether individuals would be able to integrate non-simultaneous partial visual and auditory information extrapolated into a single 360° image. According to our predictions, utilizing the auditory modality to augment the limitations of the visual system spatially will shed light on sighted participants' abilities to extend their perception beyond the natural range and perceive spatially dispersed visual information. This ability has not been previously tested using visual-to-audio SSDs. In addition, the results will provide insight into utilizing 360° audio cues to expand the normal SSD range from a 2D image to the surrounding 3D space. We also wish to test such a system's impact on shape recognition and generalization to untrained visual shapes. Finally, we discuss several practical and more theoretical neuroscience-based future directions following this approach.

Materials and methods

Participants

A total of 15 participants (6 women, aged 28.5 ± 5.8 years) with no known neurological or sensory impairments participated in the study. All participants reported normal or corrected-to-normal hearing and vision. The institutional review board (IRB) of the Reichman University approved the study. All participants were recruited *via* social media and signed an informed consent form. They were provided 40 nis per hour compensation for their participation in the study and had no prior familiarity with the EyeMusic device, the algorithm, or any other SSDs. In determining an appropriate sample size, we followed along the lines of a previous study conducted by our lab as a proof of concept introducing the novel EyeMusic algorithm, on which the current algorithm is based. The study by Abboud et al. (2014) was conducted on 12 participants.

Apparatus

The study took place in a cube-shaped soundproofed room, 408 cm (length) × 400 cm (width) × 268 cm (height) in size. A total of 72 loudspeakers were arranged along the walls in three horizontal rings at the following heights: 48, 148, and 248 cm, with an even azimuthal spacing of 15° among each adjacent pair (**Figure 1A**). Furthermore, 25 additional speakers were mounted on the ceiling in a 5 × 5 grid. All speakers in the room were measured and calibrated for spectral and delay matching. Participants were placed in the center of the room, with their heads at the height of the middle speaker ring. The center point was calibrated to a height of 148 cm (the level of the middle horizontal axis on the wall). The study was operated from a separate soundproofed control room. Interaction with the participant was carried out using a talkback microphone system and camera. The auditory stimuli were played from a local PC. Multichannel digital-to-analog conversion utilized a Dante network in combination with 13 Crown DCi 8| 300DA network-enabled amplifiers. The visual projection was performed using a sound transparent screen and a projector calibrated to fill between −45° and 45° of the front-facing

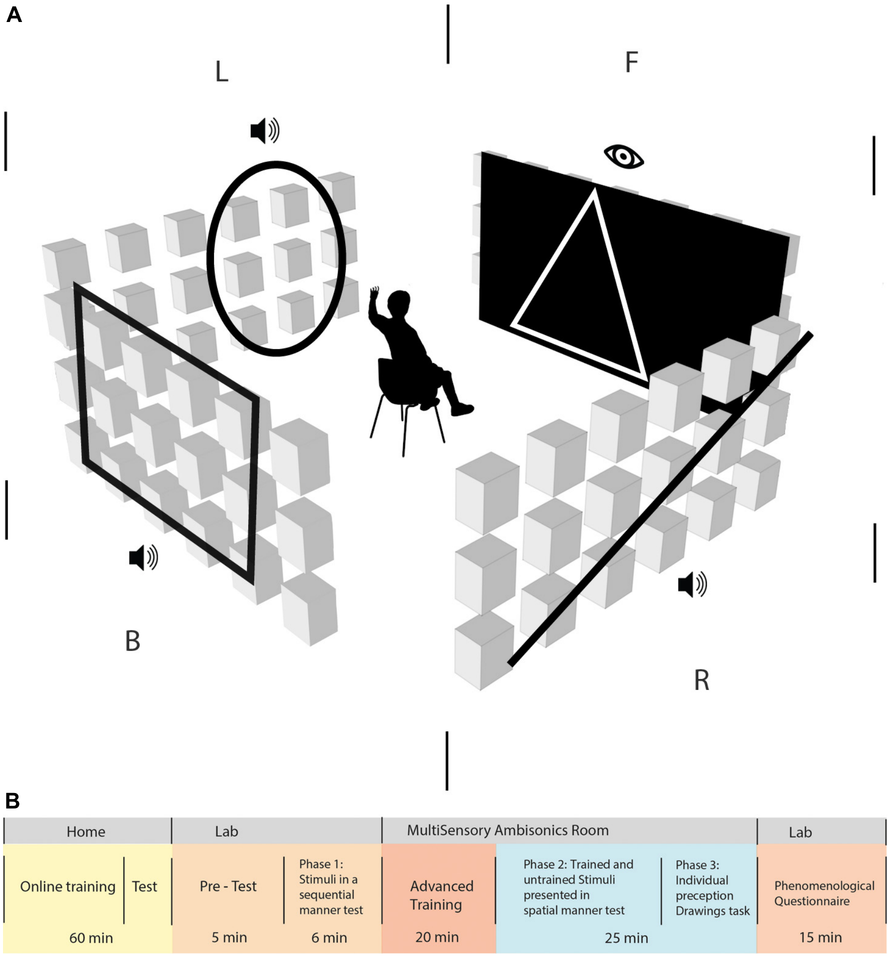


FIGURE 1
(A) Simulation of the Vision360 experiment in the multisensory room at the lab. The participants sat in the middle of the room, 2 m away from each wall, at an ear height of 1.4 m. An ambisonic system operated 21 speakers on every wall, with the corner speaker column shared among adjacent walls. F | R | B | L represents the division of space to front, right, back, and left, respectively, in an egocentric manner (in relation to the participant's body and position). Participants were required to focus on the front side of the room, where they perceived a visual stimulus and then an auditory stimulus. It appeared only on the participants' right, back, and left sides, in this order. Trained EyeMusic shapes and untrained shapes are perceived as stimuli both visually and auditorily. **(B)** Study outline. Participants went through various phases to complete the experiment: 60 min of basic training on the EyeMusic SSD at home. Following this, participants who passed the online training test at home were invited to the experiment performed at the lab and in the MultiSensory Ambisonics room. Upon the arrival of the participants to the lab, they went through a 5-min pre-test on the EyeMusic SSD material before moving on to phase 1 of the experiment. Phase 1 of the experiment was a test of stimuli presented sequentially for 6 min. Then, they moved on to the advanced training phase, where stimuli were presented to them in the ambisonics system for 20 min. Phase 2 of 25 min was a test that consisted of trained and untrained stimuli, which was presented spatially in the room. Then, they moved to phase 3, where participants were asked to draw the stimuli they perceived, and finally, they took a phenomenological questionnaire.

azimuthal angle (the entire front wall), enabling synchronized visual and auditory stimuli.

360° audio-visual transformation

Visual-to-auditory transformation of the current study, Vision360, is based on techniques of the EyeMusic technology (Abboud et al., 2014), in which pixels along the Y-axis (bottom to top) are converted to consecutive pitches along a pentatonic scale. The Vision360 algorithm takes a similar approach to vertical pixel positioning to pitch conversion while adding a spatial audio component, enabling the position of the sound source to arrive at any azimuthal position (360°) encircling the person. To maintain coherence with the original EyeMusic (in which temporality and spatial positioning are linked) and the sweep direction remains constant, in the present study, we transform the left-to-right sweep-line into a clockwise azimuthal-sweep surrounding the person. Each white-colored pixel is converted into a MIDI note pertaining to its location in the image, creating a MIDI file that can then be played back at a chosen speed. The image ratio used for all stimuli was 180 × 30 pixels (180 horizontal and 30 vertical), where the X-axis is understood to contain the full 360° ring surrounding the participant (e.g., if $X = 0$ is set to the front center of the room, then $X = 90$ will be at the back center of the room). The experiment was programmed as a Max MSP patcher, utilizing the Spat5~ library for Higher Order Ambisonics (HOA) encoding and decoding. Ambisonics is a mode of recording and reproduction based on a representation of the sound field excitation as a decomposition into spherical or circular harmonics, respectively. This achieves a physically accurate sound field reproduction restricted within a designated spot in the center of a loudspeaker array.

Stimuli

In the present study (that included only seeing individuals), all images contained white pixels with a black backdrop. All stimuli in their entirety lasted 10 s, with a proportional division of space converted to time (i.e., visual 2.5 s, audio 7.5 s). Each visual/auditory pixel was representative of ∼2° of the surrounding space. All stimuli used a pleasant-sounding timbre created using simple additive synthesis (combined sine tones based on a fundamental frequency) and its overtones (whole number multiples of the fundamental frequency) at lesser amplitudes. Playback was set to a comfortable hearing level (SPL ∼65 dB). All pitch frequencies used were between 49 and 3,135 Hz.

Vision360 stimuli

Vision360 stimuli were a combination of a visual projection followed by spatialized audio. These stimuli started as a visual projection onto the frontal 90° screen (between −45° and 45° azimuth, 40 × 30 pixels; vertical range between 30° and −28°), then the projected visuals disappeared and were followed by the remaining 270° presented through spatially moving sound. The virtual sound source moved in a clockwise manner, beginning at the 45° azimuthal point and commencing to encircle the participant fully.

Basic at-home training

Following recruitment, participants underwent approximately 60 min of home training (for study outline, see **Figure 1B**). They trained on monophonic renditions of stimuli from home, using headphones and a website platform created for the experiment. Home training included nine lessons, each including several images. Clicking under the images activated audio representations of the image generated by the EyeMusic algorithm. Each lesson was followed by a short quiz, including five multiple-choice questions, to give feedback to participants. Training consisted of simple geometric shapes such as a square, triangle, circle, horizontal/vertical/diagonal lines, arrows, simple house, happy/sad/indifferent face, and "F"/"H"/"E"/"N" letters (see **Figure 2A**). At the end of the home training, participants received a final test containing 10 Alternative Forced Choice (4AFC) questions and needed a score of 70% to pass to the subsequent phases of the experiment. Participants who passed were invited to the lab within 7–15 days after completing the home training. All participants who were invited to learn the algorithm at home succeeded to pass the test, except for one who did not complete the home training and did not take the test.

In laboratory pre-test

Once they arrived at the lab, participants received a brief explanation about the experimental space. Then they underwent a ∼5-min review of the EyeMusic algorithm. Participants then retook the final test from home training (including ten 4AFC questions) to ensure comprehension of basic EyeMusic principles learned during the online training. This test was deemed necessary because participants in the experiment arrived at the lab 5–7 days after passing the online training. Thus, it was essential to validate that the material was well remembered before moving forward to subsequent phases of the experiment conducted physically in the lab. Participants could also choose to retake this test while in the lab to ensure comfotability with the learned material. Seven participants took the test twice consecutively, the rest took it once.

Shapes in sequence: Phase 1

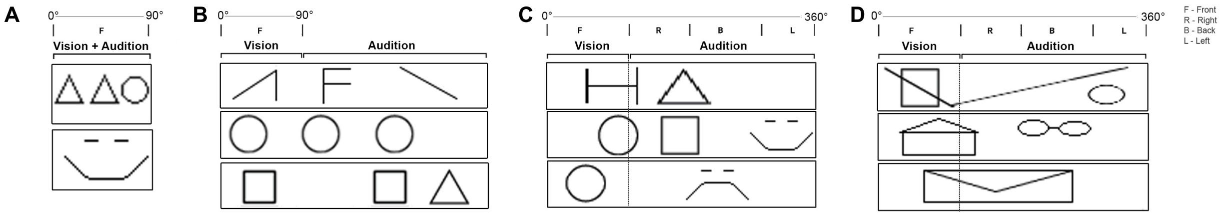
In the first experimental phase (phase 1), participants were seated in the center of the experiment room facing forward and asked to fixate their gaze on a 5 × 5 cm red square placed at the 0° azimuth and 0° elevation points. Each stimulus was composed of two to three shapes, starting with a visual shape presented on the front screen, which disappeared, followed by two separate shapes played through headphones (see **Figure 2B**). Audio given during this stage was a monophonic rendering and played equally to both ears. All shapes shown during this phase were familiar to the participants from prior training. The test included ten stimuli. Each stimulus was presented only once, starting and ending with a cue sound from the ceiling to notify the participants when a new stimulus was about to start and after it had ended. Stimuli from this phase are referred to as "shapes in sequence." Following each stimulus, participants were presented with Four-Alternative-Forced-Choice (4AFC) questions. Each possible answer displayed three shapes in chronological sequence from which they had to identify the one they were presented with (i.e., the sequence simulated the order of presentation; the left shape was always the visual shape).

Advanced training

Before the second phase of the experiment, participants underwent another training session of approximately 20 min. During this session, participants were introduced to the Vision360 transformation, then presented with 18 stimuli. The purpose of the training was to introduce the participants to the 360° audio abilities of the room and let them adjust to its immersive nature. As before, participants were asked to fixate their gaze. In this phase, some stimuli were similar to those the participants had already learned, while others were new (untrained). During training, participants were given feedback and told if their responses were correct or incorrect. If an answer was incorrect, the stimulus was repeated. Participants could ask to repeat the stimuli as many times as they needed. Performance was assessed with 4AFC questions.

360° audio representation: Phase 2

During phase 2 of the experiment, participants were tested on 22 Vision360 stimuli. Of the 22 stimuli, 11 consisted of familiar shapes, beginning with a visual shape on the front screen and then the additional 1–2 shapes presented through audio at different locations in the 360° space (see **Figure 2C**). These 11 stimuli included previously untrained shapes and consisted of long extended shapes played throughout the 360° surrounding space (see **Figure 2D**).



EyeMusic Online Training Stimuli

Stimulus in sequence

Trained stimulus in a spatial manner

Untrained stimulus in a spatial manner

FIGURE 2

Various types of stimuli used during the experiment. **(A)** Images used during the online training. Participants see the visual image and simultaneously hear the monophonic audio rendition that represents this image. **(B)** Stimulus in sequence, participants perceived in their front 90° a visual shape and afterward heard the rest of the auditory stimulus in headphones. **(C)** Stimuli with three shapes in sequence presented 360° around the participant. The front is visually projected, while the sides and back are rendered as spatial audio. **(D)** Unified audio-visual stimuli presented partially in visual and partially in audio.

These stimuli could consist of two familiar shapes or new ones being played simultaneously. Stimuli could also include shapes presented as truncated at the edge of the screen and completed in audio. All stimuli were followed by a 4AFC question depicting the 360° stimulus as one of the choices. Other possible responses contained different (yet similar) shapes to the original at similar positioning. No feedback was provided to participants following their responses. Phase 2 included a 5-min break.

Drawing 360° images: Phase 3

Finally, participants were presented with 13 additional Vision360 stimuli, which they were asked to draw in their entirety (i.e., composed of both visual and auditory segments) on a piece of paper handed to them before the task. A few examples of participants' drawings are presented in Figure 3 (for the rest of the drawings, see Supplementary material). These 13 stimuli were divided into the following three categories: category 1 included six stimuli, each containing 2–4 trained shapes placed at different locations; category 2 included two stimuli, each containing a single extended shape meant to test whether participants experienced the auditory and visual information as unified; and category 3 included five stimuli, each containing a combination of an extended shape along with smaller trained shapes in tandem. Assessment of the drawings was done by counting and rating by (1) quantity: the number of shapes drawn; (2) shape recognition: whether the shapes in themselves were correctly identified; (3) positioning: whether the shapes were placed in the proper positions; and (4) unifying audio-visual: whether the shapes broken between visual and audio were drawn connected as a single shape. Separate average scores were given for each group of stimuli.

For *category 1* (stimuli containing separated trained shapes), participants were scored in all four measures (quantity, shape recognition, positioning, and unifying audio-visual). The number of shapes drawn was compared to the number of shapes in the original stimuli. A point fraction was deduced for any shape added to or missing from the stimulus (e.g., if someone drew 2 or 4 shapes in a stimulus that had three shapes altogether, they would receive 2/3). Participants were also scored on the correctness of the shapes. Points were given for every correct shape and divided by the number of shapes (as above). Positioning scores were given based on whether participants correctly located the shapes within or between the F| R| B| A brackets. Scores were calculated in the same manner as above. Two of the five stimuli in this category had a shape broken between the visual and auditory fields. We tested whether participants recognized the shape as unified and scored with a binary rating (1 or 0). A unified shape would contain a continuous drawn line going over the "F| R" brackets).

In *category 2* (expanded single shapes), the number of shapes drawn was compared to the number of shapes perceived in the original stimuli. In shape recognition, a point fraction was subtracted if they did not unify the shape (breaking the drawn line between the "F| R" brackets). The position of these shapes was rated according to their location reflected in the brackets. Unifying audio-visual was rated in a binary rating in the same manner as in the category above.

Category 3 (combining expanded and trained shapes in tandem) was scored similarly to category 1.

The group score for each of the measures within each category was calculated by summing each participant's stimuli scores within the measure, then averaged among the group and converted to percentages (see Supplementary material for participant results).

Questionnaire

The phenomenological questionnaire was based on a questionnaire from the study by Buchs et al. (2021). It included questions regarding the perceived learning and perceived difficulty in each experiment stage and the pleasantness of the stimuli.

Verbal interview

After completing the study, we conducted a verbal interview with each participant to assess more accurately the participants' subjective experiences. They were asked to freely describe their experience, their perception of the shapes and whether or not they closed their eyes. Further questions pertained to experienced unification of shapes that began visually and continued auditorily. Finally participants were asked regarding their overall experience, beginning with the online training and commencing with the last task of the experiment.

Statistical analysis

All statistical analyses were performed using JASP (version 0.25). Wherever relevant, p-values reported in the results were corrected for multiple comparisons. All significant p-values remained significant

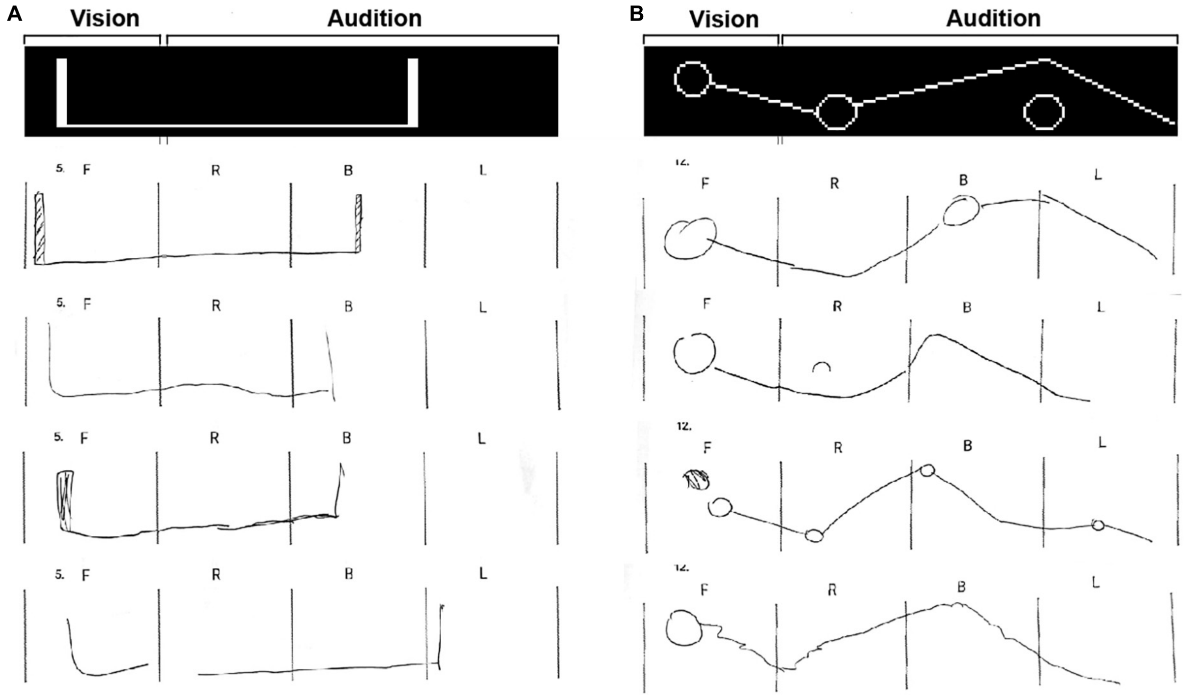


FIGURE 3

Two examples of the drawing task stimuli with 8 participants' drawings. **(A)** Example of expanded single shapes type of stimuli. The group score for the number of shapes was 85 ± 24.6%, the score for the accuracy of the shapes themselves was 83 ± 24.4%, the group score for the proper positioning of the shapes was 86.7 ± 29.7%, and the group score for the unified visual-auditory shape was 70 ± 45.5%. **(B)** Example of a stimulus from category 3 (combining expanded and trained shapes in tandem). The group score for the number of shapes was 77.7 ± 10.3%, the score for the accuracy of the shapes themselves was 43.7 ± 17%, the group score for the proper positioning of the shapes was 66 ± 14.6%, and the group score for the unified visual-auditory shape was 65.3 ± 36.6%.

after correction. To assess whether our experimental, control, trained, and generalized conditions displayed above chance level mean correct response (MCR), we performed a tailed one-sample Wilcoxon test against an alternative mean of 25%. To assess whether subjects improved between conditions, we performed a two-tailed Wilcoxon signed-rank test between conditions.

Results

Participants can successfully learn the EyeMusic algorithm using a brief online protocol. The percent of correct responses in the online test stood at 89.3 ± 5.5% (mean ± SD; **Figure 4A**), revealing a high rate of correct answers, significantly above chance level ($W(14) = 120$, $p < 0.001$).

Participants performed better in a spatial task than in a sequential task. The success rate for shapes in sequence stood at 54.6 ± 16.8% and was significantly above the chance level ($W(14) = 120$, $p_{corr} < 0.001$), showing that participants recognize stimuli composed of both visual shapes and monophonic audio-rendered fragments in sequence (**Figure 4**). When participants underwent a similar task under the same temporal condition in a 360° space, they had a correct response rate of 78.78 ± 12.2%, significantly higher than chance level ($W(14) = 120$, $p_{corr} < 0.001$) as well as higher than perceived stimuli in a sequence ($W(14) = 117$, $p_{corr} < 0.01$).

Using Vision360, participants were better at recognizing generalized stimuli than stimuli presented sequentially in monophonic rendering. We tested the participants' ability to generalize shapes received in 360°. Participants had a correct response of 82.4 ± 14.3% with a chance level of 25% ($W(14) = 120$, $p_{corr} < 0.001$). Participants were better at recognizing untrained stimuli in 360° than trained stimuli presented sequentially monophonically (during phase 1), with significantly higher results ($W(14) = 120$, $p_{corr} < 0.001$).

Participants successfully recognized both trained and untrained stimuli to a similar extent. To compare the abilities of the participants to recognize and orient the trained vs. untrained stimuli in Vision360, we performed a paired Wilcoxon test between the two conditions, which found no significant differences ($W(14) = 22$, $p = 0.349$). Meaning participants were successfully able to perform generalization.

Participants can unify shapes presented spatially, where they begin visually and end auditorily (**Figure 5**). The correct response rate for the 13 stimuli, which included shapes that started in the visual field and continued auditorily, was 76.6 ± 15.1%, significantly higher than the chance level ($W(14) = 120$, $p_{corr} < 0.001$). This finding indicates that participants can correctly unify shapes composed of visual and auditory components perceived in 360°.

Phase 3 consisted of 13 drawing tasks. In the drawing task, participants were asked to draw what they perceived both visually and auditorily. A few examples of different participants' drawings are presented in **Figure 3** (for the other participants' drawings, see **Supplementary material**).

Category 1 (stimuli containing 2–4 separated trained shapes), the group-averaged accuracy of perceiving the number of shapes was 85.6 ± 13.2%, for shape recognition was 58 ± 16.2%, for proper

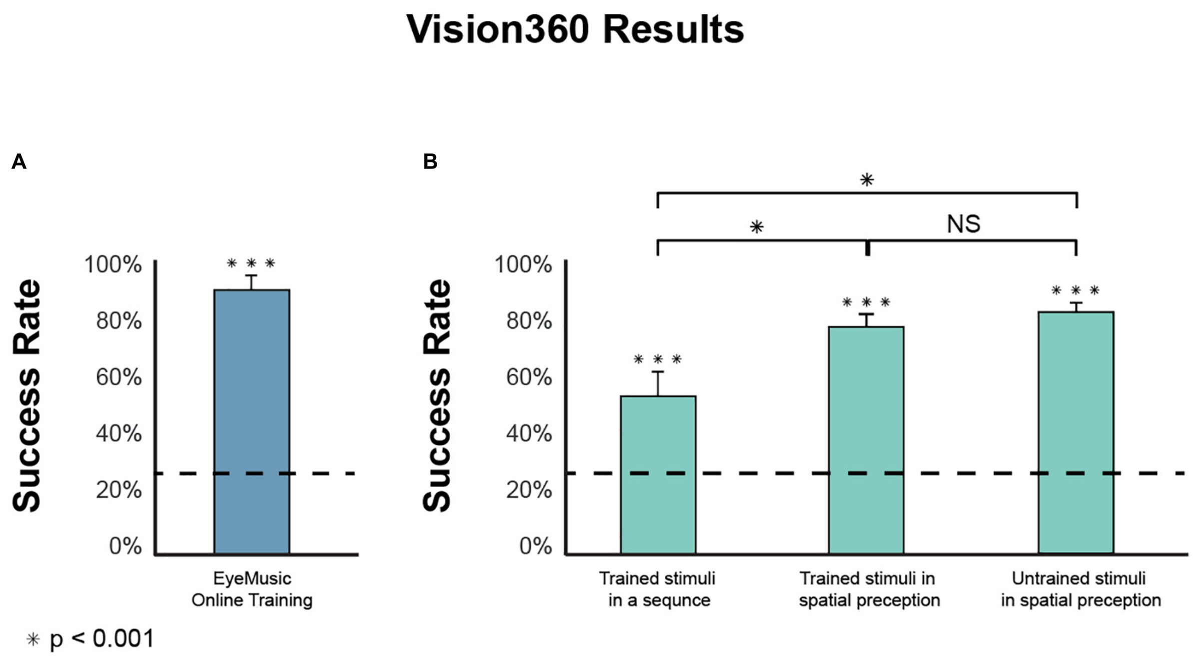


FIGURE 4

Experiments results. **(A)** Participants in the EyeMusic online training had a statistically significant success rate in the final test of 89.3 ± 5.5% [mean correct response ± SD; **(A)** bars indicate the standard error; dashed line indicates chance level]. **(B)** Experimental phase tasks' results are divided into three categories: Recognition of stimulus in a sequence (54.6 ± 16.8%), the success rate of spatially perceiving trained shapes (78.8 ± 12.2%), and recognition of untrained shapes (Generalization) perceived spatially (82.4 ± 14.3%). There was no significant difference between the trained and generalized conditions (W(14) = 22, $p = 0.349$). However, between the trained sequentially presented stimulus and the trained stimuli presented spatially, there is a significant difference (W(14) = 117, $p_{corr} < 0.01$), as well as between the stimuli in a sequence compared to the Generalization of stimuli presented spatially (W(14) = 120, $p_{corr} < 0.001$). ***Means significantly above chance. *Means significantly different between two conditions. NS, not significant.

positioning was 78 ± 13.4%, and for unifying audio-visual was 46.7 ± 41.4%.

Category 2 (expanded single shapes), the group-averaged accuracy of perceiving the correct quantity of shapes was 85 ± 24.6%, for shape recognition was 83 ± 24.4%, for proper positioning was 86.7 ± 29.7%, and for unifying audio-visual was 70 ± 45.5%.

Category 3 (combining expanded and trained shapes in tandem), the group-averaged accuracy of perceiving the correct quantity of shapes was 77.7 ± 10.3%, for shape recognition was 43.7 ± 17%, for proper positioning was 66 ± 14.6%, and for unifying audio-visual was 65.3 ± 36.6%.

Verbal interviews

In the verbal interviews, 8 out of 15 participants stated that spatial information helped them recognize and remember the shapes. Participant #12: *"The different locations of the shapes has helped me to remember them, whether a specific shape had appeared from the right or the left"*; Participant #3: *"I imagined the shapes in a way that they would immerse me around my body, then I performed some kind of flattening of the space around me to a strip. After I drew it, I would rethink it and correct the locations if needed, using my memory"*; Participant #9: *"I paid attention to the size of the room; I heard it on the left and not only in the back, so when I wasn't sure what shape it was I used the sides, it was helpful."* Out of 15 participants, 14 reported experiencing the shapes passing between visual and auditory as intuitively unified. Participant #10: "*It depends on the location of the sound. If the visual shape was cut by the end of the screen on my right, and that's where exactly the sound had appeared right after, I understood it is connected*"; participant #3: "*I think that every time I saw the visual shape halved by the end of the screen to my right, I expected that it would be completed and continued with some kind of sound.*" Out of 15 participants, 7 indicated using their index finger to trace the auditorily received visual shapes through the air. Most of them stated they did it to help them recognize the perceived shapes. Participant #14: *"I drew with my fingers in order to recognize the shapes"*; Participant #15: "*I used my fingers to physically draw and imagine the shapes, also sometimes I did close my eyes to imagine the shapes.*" Out of 15 participants, 7 closed their eyes while experiencing some of the auditory cues, stating it helped them focus on recognizing the shapes when experiencing the auditory part of the stimuli. During phase 2, participants said they were replaying melodic memory and recalling mental imagery to answer the different phases. Participant #5: *"I tried to neutralize my visual sense, and I felt it strengthened my auditory sense. When I closed my eyes, it helped me imagine what I heard."*

Discussion

The current study investigated the role of auditory spatial perception in recognizing visual geometric shapes presented in a 360° space. To achieve this, participants needed to combine visual information with auditory information conveyed through a visual-to-auditory SSD. We asked participants to detect and orient shapes by reporting their egocentric spatial location (Front/Right/Back/Left).

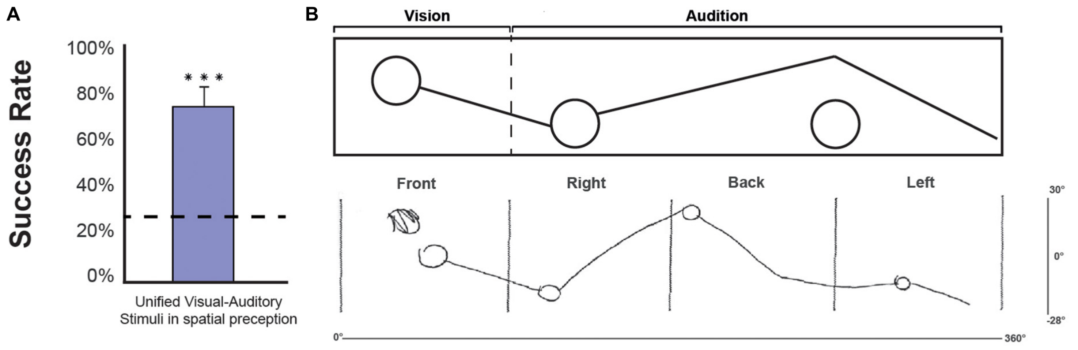


FIGURE 5
(A) Group result for shapes that began in the visual field and continued auditorily. We performed a one-tailed one-sample Wilcoxon test against chance. The correct response rate was 76.6 ± 15.1%, significantly higher than the chance level ($W(14) = 120$, $p < 0.001$) (bar indicates the standard error; dashed line indicates chance level). **(B)** An example of a full visual-auditory stimulus as processed in the Vision360 application. The "vision" section of the stimulus is perceived by the participants in the front, and the "Audition" section is perceived by the participants auditorily (starting from left to right in relation to the participants' location). Underneath the stimulus, an example of a drawing by participant number 9, taken from the drawing phase of the experiment is shown. Front | right | back | left are the space expressions standardizing the division of space for the participants according to their egocentric position in space. The x-axis of 0°–360° represents the horizontal coverage of the stimulus in space, and the y-axis of 30°–28° represents the vertical coverage of the stimulus in space. ***Means significantly above chance.

Our findings indicate that sighted participants can indeed process spatial information starting in their visual field (frontal 90°) and continuing in their auditory field (the remaining 270°) to create a unified image. In addition, this study replicated the results of a previous study conducted by Buchs et al. (2021), showing the efficacy of online training for visual-to-auditory SSDs. We show that subjects could even draw the stimuli within a short period of time, to some success when performing a task demanding the conversion of the entire surrounding image into a 2D visual rendering of the space (combining the back, front, and both sides onto a single 2D plane). To the best of our knowledge, this is the first time participants demonstrated projection of the back 3D space received through audio into a 2D drawn visual rendering.

Our 3D-rendered SSD employs insights derived from the growing body of knowledge on sensory substitution. The field of sensory substitution owes its beginnings to the domain of sensory rehabilitation, with research initially being conducted on conveying visual information to the blind through an alternate sense (for a review, see Maidenbaum et al., 2014). In recent years, this field has burgeoned, with several sensory substitution systems and algorithms currently being developed for various research aims. Most of these are based on the substitution of visual information through the auditory or tactile systems (Bach-y-Rita et al., 1969; Bach-y-Rita, 1983, 2004; Ptito et al., 2005; Chebat et al., 2007) and the substitution of auditory information through the tactile system (Cieśla et al., 2019, 2022), among others. Recent studies in our lab, both in the sighted (Netzer et al., 2021) and in the blind (Maimon et al., 2023), have also tested the ability to extend/augment visual-spatial perception (in both the front and the back) using auditory cues.

This study follows along these lines to provide a proof of concept for the unique system and algorithm, which builds on our prior work but takes it one step further, with the aim of pushing the limits of our current senses by providing complementary information simultaneously through other modalities. Other studies exploring such sensory enhancement include Nardini (2021) and Negen et al. (2021), who looked into integrating distance perception with an echolocation auditory type cue. Negen et al. (2021) indicated that sensory integration can become automatic, a finding with significant implications. Recently, Witzel et al. (2022) published a study exploring the automaticity of novel perceptual experiences by employing a sensory augmentation device for perceiving the north direction. These studies further support the subjective reports presented in case studies that indicate acquired automaticity and transparency following extensive use of sensory substitution devices (Ward and Meijer, 2010; Maimon et al., 2022).

As we have demonstrated in this study, subjects can perceive a shape (an abstract concept associated with the visual modality) as a combination of visual and auditory information. Our findings during the 4AFC tasks indicate that participants could use the spatial cues to heighten their success compared with monophonic renderings of the algorithm. They further suggest that fundamental advantages of the original EyeMusic, such as generalizability, remained possible and intuitive when making a move to 3D. We believe these findings are related to the fact that combining information from these modalities in our surrounding space takes place constantly, and indeed the ability to localize audio is thought to be constantly calibrated visually (Knudsen and Knudsen, 1989; Zwiers et al., 2001; Gori et al., 2014).

In our prior research, tactile inputs have also been used to show spatial awareness (Yizhar et al., 2021; Snir et al., under review). Nevertheless, in the present study, visuals and audio appear as temporally and spatially completing one another, with no overlap. The fact that this can be unified into a single visual percept strengthens the claim that spatial perception is multisensory in nature and can be recalled as such (Quak et al., 2015). This is reinforced when considering participants' interview responses, where some indicated that the added spatial component created a more vivid and memorable experience. They also recall using spatial cues toward the reconstruction of the entire stimulus.

Further studies are warranted to see whether this could impact memory abilities in such tasks. We also believe participant accounts of using their finger to draw the stimuli through the air may provide further qualitative evidence toward a multisensory understanding of spatial information, and indeed, in this case, recruiting motor actions for the task (Clark, 2003). On the other hand, the fact that nearly half of the participants closed their eyes to better concentrate on the audio may also indicate the dominance of vision over the auditory system, as seen in various studies (Colavita, 1974; Hutmacher, 2019). Lowered activation of the auditory cortex during visual memory tasks should also be taken into consideration here (Azulay et al., 2009). Although various sensory areas have already been shown to be influenced by more than one sense (for a review, see Heimler et al., 2015 and Amedi et al., 2017), a single area representing spatial understanding has yet to be found. We believe the Vision360 technology may assist in further advancing research in this direction.

To further understand the perception of the 360° images, we employed a drawing task where we asked participants to reproduce their spatial experience onto a steady 2D plane. Although it is known in cognitive studies that spatial expressions involve some degree of ambiguity (Imai et al., 1999), in the current experiment, we utilized the structure of the experimental space (see the "Methods" section) to make clear borders between each egocentric spatial expression according to each face of the room, while maintaining the original fixed temporal sweep, in line with the EyeMusic logic. Participants made the 3D to 2D conversion intuitively and with no added training. This may be partly inherited from the scoring system, in which a missing will also lose points for shape recognition and positioning by default. Nevertheless, drawing of extra shapes would have the reverse effect, losing points for count while having no impact on the recognition and positioning scores.

Furthermore, recognition of EyeMusic stimuli may still be more challenging than spatial positioning of visuals and sound or accounting of spatially distributed objects because our healthy participants have had an entire lifetime to learn such multisensory tasks, as opposed to the conversion of temporal and auditory frequency information toward shape recognition, which they only had about an hour and a half of experience with altogether. It may nevertheless indicate multisensory spatial information as being more easily geared toward orientation than toward sensory particular information such as shape recognition. The fact that unification of the auditory and the visual information takes place at least some of the time in the majority of participants further strengthens this possibility, considering the continuous motion was presented as continuous in spatial orientation among both senses. A previous study that had a greater variety of visual shapes experienced as auditory cues tested the human ability to perceive biological movement through friction sounds produced by the action of drawing; similarly to our study, the drawings were of geometric shapes and showed the intuitive connection between kinematic movements and auditory cues (Thoret et al., 2014). This again demonstrates the multisensory connection between vision, audition, and the motor system (Jeannerod, 1995; King et al., 2009).

We use the task of drawing as a method of gauging recognition and orientation of the stimuli. The correspondence between the information provided in vision and audition and the 2D image drawn by the participants showed clear similarities. Drawings are commonly used in contemporary music to either describe or create music (Thiebaut et al., 2008). The use of drawings has also led to some interesting applications, including the development of new sonification strategies (Andersen and Zhai, 2008). Hence, drawing seemed to be a natural way of describing the motion evoked by sounds and controlling perceptually relevant attributes. Research on the blind, including a case study conducted by our lab on a blind artist, indicates an overlap between areas in the brain involved in vision and mental imagery (Amedi et al., 2008). As people can create a coherent image of their combined visual and auditory experience, it would be interesting to explore these mechanisms in the brain and see their overlap (or lack thereof). Further investigation could warrant testing for enhanced connectivity following training on Vision360, something that has been shown to occur with gradually decaying vision in adults (Sabbah et al., 2016).

Future research directions will use functional MRI to explore the possibility of novel topographies in the brain following training with sensory augmentation systems such as the Vision360 utilized in the present study. Initial research in our lab supports this idea, showing the emergence of new topographic maps following sensory substitution training and use, specifically concerning audio-rendered musical fragments similar to those used in this experiment (Hofstetter et al., 2021). Such findings may have implications for classic concepts such as the division of the brain into senses and Hubel and Wiesel's theory of critical periods (Wiesel and Hubel, 1965). The fact that the natural perceptual capabilities can be expanded through integrating two senses well into adulthood may strengthen interpretations that call into question these two seminal theories. Yet, this matter warrants further investigation in future studies. The findings of such studies may suggest, on the one hand, that the critical periods are not as strict as has previously been accepted, and on the other that the brain is perhaps divided by tasks rather than senses, strengthening the task-specific sensory independent theory of brain development and organization (Heimler et al., 2015; Amedi et al., 2017; Heimler and Amedi, 2020). This study adds to the cumulative evidence from many studies across the last couple of decades, specifically employing sensory substitution, and perceptual cross-modal learning, the findings of which suggest that the aforementioned theories warrant revision, including the metamodal theory of brain organization (Pascual-Leone and Hamilton, 2001; Cecchetti et al., 2016) and the supramodal interpretation (Kupers and Ptito, 2011). We speculate that further findings into novel topographies in the brain resulting from training on such sensory augmentation systems would further promote this paradigm shift, and we believe our system could be employed for insights into this matter.

Data availability statement

The datasets generated and analyzed during the current study are available in the Open Science Framework repository https://osf.io/3ntx4/.

Ethics statement

The studies involving human participants were reviewed and approved by the Reichman University Institutional Review Board (IRB). The patients/participants provided their written informed consent to participate in this study.

Author contributions

SS: writing—original draft, writing—review and editing, project administration, methodology, conceptualization, visualization, investigation, and formal analysis. AS: writing—original draft, writing—review and editing, methodology, conceptualization, visualization, and formal analysis. AM: writing—original draft, writing—review and editing, conceptualization, project administration, and supervision. OY: writing—original draft, writing—review and editing, methodology, conceptualization, and supervision. SH: writing—original draft, methodology, conceptualization, and investigation. KP: methodology, conceptualization, investigation, and formal analysis. AA: writing—original draft, writing—review and editing, project administration, supervision, resources, conceptualization, investigation, methodology, and funding acquisition. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by an ERC Consolidator Grant (773121 NovelExperiSense) and a Horizon GuestXR (101017884) grant — both to AA.

Acknowledgments

We would like to thank Nitsan Lavie for his contribution to preparing the online home training.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2023.1058617/full#supplementary-material

References

Abboud, S., Hanassy, S., Levy-Tzedek, S., Maidenbaum, S., and Amedi, A. (2014). EyeMusic: Introducing a "visual" colorful experience for the blind using auditory sensory substitution. *Restor. Neurol. Neurosci.* 32, 247–257. doi: 10.3233/RNN-130338

Alais, D., Newell, F., and Mamassian, P. (2010). Multisensory processing in review: From physiology to behaviour. *Seeing Perceiv.* 23, 3–38. doi: 10.1163/187847510X488603

Amedi, A., Hofstetter, S., Maidenbaum, S., and Heimler, B. (2017). Task selectivity as a comprehensive principle for brain organization. *Trend. Cogn. Sci.* 21, 307–310. doi: 10.1016/j.tics.2017.03.007

Amedi, A., Merabet, L. B., Camprodon, J., Bermpohl, F., Fox, S., Ronen, I., et al. (2008). Neural and behavioral correlates of drawing in an early blind painter: A case study. *Brain Res.* 1242, 252–262. doi: 10.1016/j.brainres.2008.07.088

Amedi, A., Stern, W. M., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., et al. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nat. Neurosci.* 10, 687–689. doi: 10.1038/nn1912

Andersen, T. H., and Zhai, S. (2008). "Writing with music" Exploring the use of auditory feedback in gesture interfaces. *ACM Trans. Appl. Percept.* 7, 1–24. doi: 10.1145/1773965.1773968

Azulay, H., Striem, E., and Amedi, A. (2009). Negative BOLD in sensory cortices during verbal memory: a component in generating internal representations? *Brain Topogr.* 21, 221–231. doi: 10.1007/s10548-009-0089-2

Bach-y-Rita, P. (1983). Tactile vision substitution: Past and future. *Int. J. Neurosci.* 19, 29–36. doi: 10.3109/00207458309148643

Bach-y-Rita, P., Collins, C. C., Saunders, F. A., White, B., and Scadden, L. (1969). Vision substitution by tactile image projection. *Nature* 221, 963–964. doi: 10.1038/221963a0

Bach-y-Rita, P. (2004). Tactile sensory substitution studies. *Ann. N. Y. Acad. Sci.* 1013, 83–91. doi: 10.1196/annals.1305.006

Bizley, J. K., and Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nat. Rev. Neurosci.* 14, 693–707. doi: 10.1038/nrn3565

Boyer, E. O., Vandervoorde, L., Bevilacqua, F., and Hanneton, S. (2015). "Touching sounds: audio virtual surfaces," in *Proceedings of the 2015 IEEE 2nd VR Workshop on Sonic Interactions for virtual Environments (SIVE)*, (Piscataway, NJ: IEEE), 1–5. doi: 10.1109/SIVE.2015.7361291

Bruns, P. (2019). The ventriloquist illusion as a tool to study multisensory processing: An update. *Front. Int. Neurosci.* 13:51. doi: 10.3389/fnint.2019.00051

Buchs, G., Haimler, B., Kerem, M., Maidenbaum, S., Braun, L., and Amedi, A. (2021). A self-training program for sensory substitution devices. *PLoS One* 16:e0250281. doi: 10.1371/journal.pone.0250281

Carello, C., Anderson, K. L., and Kunkler-Peck, A. J. (1998). Perception of object length by sound. *Psychol. Sci.* 9, 211–214. doi: 10.1111/1467-9280.00040

Casati, R., and Varzi, A. C. (1996). The structure of spatial localization. *Philos. Stud.* 82, 205–239. doi: 10.1007/BF00364776

Cecchetti, L., Kupers, R., Ptito, M., Pietrini, P., and Ricciardi, E. (2016). Are supramodality and cross-modal plasticity the yin and yang of brain development? From blindness to rehabilitation. *Front. Syst. Neurosci.* 10:89. doi: 10.3389/fnsys.2016.00089

Chebat, D. R., Rainville, C., Kupers, R., and Ptito, M. (2007). Tactile–'visual'acuity of the tongue in early blind individuals. *Neuroreport* 18, 1901–1904. doi: 10.1097/WNR.0b013e3282f2a63

Cieśla, K., Wolak, T., Lorens, A., Heimler, B., Skarżyński, H., and Amedi, A. (2019). Immediate improvement of speech-in-noise perception through multisensory stimulation via an auditory to tactile sensory substitution. *Restor. Neurol. Neurosci.* 37, 155–166. doi: 10.3233/RNN-190898

Cieśla, K., Wolak, T., Lorens, A., Mentzel, M., Skarżyński, H., and Amedi, A. (2022). Effects of training and using an audio-tactile sensory substitution device on speech-in-noise understanding. *Sci. Rep.* 12, 1–16. doi: 10.1038/s41598-022-06855-8

Clark, H. H. (2003). "Pointing and placing," in *Pointing: Where language, culture, and cognition meet*, ed. S. Kita (Mahwah, NJ: Lawrence Erlbaum Associates Publishers.).

Colavita, F. B. (1974). Human sensory dominance. *Percept. Psychophys.* 16, 409–412. doi: 10.3758/BF03203962

Collignon, O., Voss, P., Lassonde, M., and Lepore, F. (2009). Cross-modal plasticity for the spatial processing of sounds in visually deprived subjects. *Exp. Brain Res.* 192, 343–358. doi: 10.1007/s00221-008-1553-z

Erdogan, G., and Jacobs, R. A. (2017). Visual shape perception as Bayesian inference of 3D object-centered shape representations. *Psychol. Rev.* 124:740. doi: 10.1037/rev0000086

Ernst, M. O. (2008). Multisensory integration: A late bloomer. *Curr. Biol.* 18, R519–21. doi: 10.1016/j.cub.2008.05.002

Gori, M., Del Viva, M., Sandini, G., and Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Curr. Biol.* 18, 694–698. doi: 10.1016/j.cub.2008.04.036

Gori, M., Vercillo, T., Sandini, G., and Burr, D. (2014). Tactile feedback improves auditory spatial localization. *Front. Psychol.* 5:1121. doi: 10.3389/fpsyg.2014.01121

Heimler, B., and Amedi, A. (2020). Are critical periods reversible in the adult brain? Insights on cortical specializations based on sensory deprivation studies. *Neurosci. Biobehav. Rev.* 116, 494–507. doi: 10.1016/j.neubiorev.2020.06.034

Heimler, B., Striem-Amit, E., and Amedi, A. (2015). Origins of task-specific sensory-independent organization in the visual and auditory brain: Neuroscience evidence, open questions and clinical implications. *Curr. Opin. Neurobiol.* 35, 169–177. doi: 10.1016/j.conb.2015.09.001

Hertz, U., and Amedi, A. (2015). Flexibility and stability in sensory processing revealed using visual-to-auditory sensory substitution. *Cereb. Cortex* 25, 2049–2064. doi: 10.1093/cercor/bhu010

Hofstetter, S., Zuiderbaan, W., Heimler, B., Dumoulin, S. O., and Amedi, A. (2021). Topographic maps and neural tuning for sensory substitution dimensions learned in adulthood in a congenital blind subject. *Neuroimage* 235:118029. doi: 10.1016/j.neuroimage.2021.118029

Hutmacher, F. (2019). Why is there so much more research on vision than on any other sensory modality? *Front. Psychol.* 10:2246. doi: 10.3389/fpsyg.2019.02246

Imai, M., Nakanishi, T., Miyashita, H., Kidachi, Y., and Ishizaki, S. (1999). The meanings of front/back/left/right. *Cogn. Stud.* 6, 207–225.

Jeannerod, M. (1995). Mental imagery in the motor context. *Neuropsychologia* 33, 1419–1432. doi: 10.1016/0028-3932(95)00073-C

King, B. R., Kagerer, F. A., Contreras-Vidal, J. L., and Clark, J. E. (2009). Evidence for multisensory spatial-to-motor transformations in aiming movements of children. *J. Neurophysiol.* 101, 315–322. doi: 10.1152/jn.90781.2008

Knudsen, E. I., and Knudsen, P. F. (1989). Vision calibrates sound localization in developing barn owls. *J. Neurosci.* 9, 3306–3313. doi: 10.1523/JNEUROSCI.09-09-03306.1989

Kupers, R., and Ptito, M. (2011). Insights from darkness: What the study of blindness has taught us about brain structure and function. *Prog. Brain Res.* 192, 17–31. doi: 10.1016/B978-0-444-53355-5.00002-6

Maidenbaum, S., Abboud, S., and Amedi, A. (2014). Sensory substitution: Closing the gap between basic research and widespread practical visual rehabilitation. *Neurosci. Biobehav. Rev.* 41, 3–15. doi: 10.1016/j.neubiorev.2013.11.007

Maimon, A., Wald, I. Y., Ben Oz, M., Codron, S., Netzer, O., Heimler, B., et al. (2023). The topo-speech sensory substitution system as a method of conveying spatial information to the blind and vision impaired. *Front. Hum. Neurosci.* 16:1058093. doi: 10.3389/fnhum.2022.105

Maimon, A., Yizhar, O., Buchs, G., Heimler, B., and Amedi, A. (2022). A case study in phenomenology of visual experience with retinal prosthesis versus visual-to-auditory sensory substitution. *Neuropsychologia* 173:108305. doi: 10.1016/j.neuropsychologia.2022.108305

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0

Millodot, M. (2014). *Dictionary of optometry and visual science e-book*. Amsterdam: Elsevier Health Sciences.

Milner, P. M. (1974). A model for visual shape recognition. *Psychol. Rev.* 81:521. doi: 10.1037/h0037149

Nardini, M. (2021). Merging familiar and new senses to perceive and act in space. *Cogn. Process.* 22, 69–75. doi: 10.1007/s10339-021-01052-3

Negen, J., Bird, L. A., Slater, H., Thaler, L., and Nardini, M. (2021). A new sensory skill shows automaticity and integration features in multisensory interactions. *Biorxiv* [Preprint]. doi: 10.1101/2021.01.05.425430

Netzer, O., Heimler, B., Shur, A., Behor, T., and Amedi, A. (2021). Backward spatial perception can be augmented through a novel visual-to-auditory sensory substitution algorithm. *Sci. Rep.* 11, 1–14. doi: 10.1038/s41598-021-88595-9

Pascual-Leone, A., and Hamilton, R. (2001). The metamodal organization of the brain. *Prog. Brain Res.* 134, 427–445. doi: 10.1016/S0079-6123(01)34028-1

Peelen, M. V., and Kastner, S. (2014). Attention in the real world: Toward understanding its neural basis. *Trends Cogn. Sci.* 18, 242–250. doi: 10.1016/j.tics.2014.02.004

Pietrini, P., Furey, M. L., Ricciardi, E., Gobbini, M. I, Wu, W. H. C., Cohen, L., et al. (2004). Beyond sensory images: Object-based representation in the human ventral pathway. *Proc. Natl. Acad. Sci. U.S.A.* 101, 5658–5663. doi: 10.1073/pnas.0400707101

Ptito, M., Moesgaard, S. M., Gjedde, A., and Kupers, R. (2005). Cross-modal plasticity revealed by electrotactile stimulation of the tongue in the congenitally blind. *Brain* 128, 606–614. doi: 10.1093/brain/awh380

Quak, M., London, R. E., and Talsma, D. (2015). A multisensory perspective of working memory. *Front. Hum. Neurosci.* 9:197. doi: 10.3389/fnhum.2015.00197

Sabbah, N., Authié, C. N., Sanda, N., Mohand-Saïd, S., Sahel, J. A., Safran, A. B., et al. (2016). Increased functional connectivity between language and visually deprived areas in late and partial blindness. *Neuroimage* 136, 162–173. doi: 10.1016/j.neuroimage.2016.04.056

Shapiro, A., Lu, Z. L., Huang, C. B., Knight, E., and Ennis, R. (2010). Transitions between central and peripheral vision create spatial/temporal distortions: A hypothesis concerning the perceived break of the curveball. *PLoS One* 5:e13296. doi: 10.1371/journal.pone.0013296

Strasburger, H. (2020). Seven myths on crowding and peripheral vision. *i-Perception* 11:2041669520913052. doi: 10.1177/2041669520913052

Striem-Amit, E., Cohen, L., Dehaene, S., and Amedi, A. (2012). Reading with sounds: sensory substitution selectively activates the visual word form area in the blind. *Neuron* 76, 640–52. doi: 10.1016/j.neuron.2012.08.026

Thiebaut, J. B., Healey, P. G., and Bryan-Kinns, N. (2008). "Drawing electroacoustic music," in *Proceedings of the 2008 international computer music conference, ICMC 2008*, Belfast.

Thoret, E., Aramaki, M., Kronland-Martinet, R., Velay, J. L., and Ystad, S. (2014). From sound to shape: Auditory perception of drawing movements. *J. Exp. Psychol. Hum. Percept. Perform.* 40:983. doi: 10.1037/a0035441

Traquair, H. M. (1938). *An introduction to clinical perimetry, chpt. 1*. London: Henry Kimpton, 4–5.

Ward, J., and Meijer, P. (2010). Visual experiences in the blind induced by an auditory sensory substitution device. *Conscious. Cogn.* 19, 492–500. doi: 10.1016/j.concog.2009.10.006

Wiesel, T. N., and Hubel, D. H. (1965). Comparison of the effects of unilateral and bilateral eye closure on cortical unit responses in kittens. *J. Neurophysiol.* 28, 1029–1040. doi: 10.1152/jn.1965.28.6.1029

Witzel, C., Lübbert, A., O'Regan, J. K., Hanneton, S., and Schumann, F. (2022). Can perception be extended to a "feel of north"? Tests of automaticity with the NaviEar. *Adapt. Behav*. doi: 10.1177/10597123221130235 [Epub ahead of print].

Yizhar, O., Buchs, G., Heimler, B., Friedman, D., and Amedi, A. (2021). Applying a novel visual-to-touch sensory substitution for studying tactile reference frames. *Sci. Rep.* 11, 1–8. doi: 10.1038/s41598-021-90132-7

Zwiers, M. P., Van Opstal, A. J., and Cruysberg, J. R. M. (2001). A spatial hearing deficit in early-blind humans. *J. Neurosci.* 21, RC142. doi: 10.1523/JNEUROSCI.21-09-j0002.2001

frontiers | Frontiers in Human Neuroscience

TYPE Brief Research Report
PUBLISHED 09 March 2023
DOI 10.3389/fnhum.2023.1026196

OPEN ACCESS

EDITED BY
Monica Gori,
Italian Institute of Technology (IIT), Italy

REVIEWED BY
Ke Ma,
Southwest University, China
Nicolas Deltort,
UMR 5287 Institut de Neurosciences Cognitives et Intégratives d'Aquitaine (INCIA), France

*CORRESPONDENCE
Stefania La Rocca
✉ stefania.larocca@unimib.it

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

SPECIALTY SECTION
This article was submitted to Sensory Neuroscience, a section of the journal Frontiers in Human Neuroscience

RECEIVED 23 August 2022
ACCEPTED 22 February 2023
PUBLISHED 09 March 2023

CITATION
La Rocca S, Gobbo S, Tosi G, Fiora E and Daini R (2023) Look at me now! Enfacement illusion over computer-generated faces. *Front. Hum. Neurosci.* 17:1026196. doi: 10.3389/fnhum.2023.1026196

Look at me now! Enfacement illusion over computer-generated faces

Stefania La Rocca[1,2*†], Silvia Gobbo[1,2†], Giorgia Tosi[2,3], Elisa Fiora[1] and Roberta Daini[1,2]

[1]Department of Psychology, University of Milano–Bicocca, Milan, Italy, [2]MiBTec–Mind and Behavior Technological Center, University of Milano–Bicocca, Milan, Italy, [3]Department of History, Society and Human Studies, University of Salento, Lecce, Italy

According to embodied cognition research, one's bodily self-perception can be illusory and temporarily shifted toward an external body. Similarly, the so-called "enfacement illusion" induced with a synchronous multisensory stimulation over the self-face and an external face can result in implicit and explicit changes in the bodily self. The present study aimed to verify (i) the possibility of eliciting an enfacement illusion over computer-generated faces and (ii) which multisensory stimulation condition was more effective. A total of 23 participants were asked to look at a gender-matched avatar in three synchronous experimental conditions and three asynchronous control conditions (one for each stimulation: visuotactile, visuomotor, and simple exposure). After each condition, participants were asked to complete a questionnaire assessing both the embodiment and the enfacement sensations to address different facets of the illusion. Results suggest a stronger effect of synchronous vs. asynchronous stimulation, and the difference was more pronounced for the embodiment items of the questionnaire. We also found a greater effect of visuotactile and visuomotor stimulations as compared to the simple exposure condition. These findings support the enfacement illusion as a new paradigm to investigate the ownership of different face identities and the specific role of visuotactile and visuomotor stimulations with virtual reality stimuli.

KEYWORDS

enfacement, face processing, computer-generated faces, embodiment, visual perception

1. Introduction

The self-face is recognized as a special stimulus for our face-processing system, as shown by both behavioral and neuroimaging studies (Devue and Brédart, 2011; Bortolon and Raffard, 2018; Alzueta et al., 2020). Notably, an advantage of one's face has been observed in face processing (i.e., self-face advantage; Sugiura et al., 2005; Ma and Han, 2010). This effect has been documented even in participants with difficulties recognizing faces, namely congenital prosopagnosics (Malaspina et al., 2018).

Thus, causing participants to identify with presented faces could, in principle extend the self-face advantage to faces other than one's own. This can lead to a new range of possibilities in the domain of face processing research.

A temporary change in one's self-representation can be induced by balancing multisensory information (Tsakiris, 2010). This change has been observed both with bodies (Lenggenhager et al., 2007) and faces (Tsakiris, 2008). The experimental procedures addressing the body are named body illusions and refer to the embodiment effect, which is the experience of ownership over a fake body (or parts of it). Crucially, when faces are concerned, this effect is called "Enfacement illusion"(Sforza et al., 2010).

Different methods have been proposed to induce the enfacement illusion, namely visuotactile, visuomotor, and visuotactile-motor stimulations (Porciello et al., 2018). In the visuotactile stimulation, participants look at another face in front of them while both their face and the other one are touched by a stick (Tsakiris, 2008; Sforza et al., 2010; Tajadura-Jiménez et al., 2012a,b). The touch can be synchronous with the viewed face in terms of timing and location (synchronous condition), or asynchronous (asynchronous condition). The latter usually serves as a control condition because it seems not to induce the illusion (Porciello et al., 2018). In particular, studies suggest that temporal synchrony is more important than spatial synchrony in inducing the effect (Apps et al., 2015). The visuomotor stimulation consists of participants viewing a face in front of them while being instructed to produce head movements. Movements can be synchronized with the viewed face (synchronous condition) or not (asynchronous condition). Again, the asynchronous condition usually serves as a control (Serino et al., 2015). Active movements (i.e., movements that are controlled by the participant) are suggested to be more effective than passive movements (i.e., manipulated by the experimenter) in eliciting the embodiment effect in a classical rubber hand illusion paradigm (Dummer et al., 2009). However, the effect observed by Dummer et al. (2009) was only marginally significant; moreover, another study did not find significant differences between active and passive movements (Kalckert and Ehrsson, 2014). Therefore, it is not completely clear whether movement needs to be active to elicit embodiment or enfacement. The third type of stimulation described in the literature concerns visuotactile-motor stimulation: participants perceive a touch resulting from a movement generated by themselves (Tajadura-Jiménez et al., 2013), which can be synchronous or asynchronous with the observed face. To the best of our knowledge, in enfacement paradigms the difference between active and passive stimulation in pure visuomotor condition has not been investigated. As described in Dummer et al. (2009) for embodiment, active stimulation refers to a movement elicited by the participant while passive stimulation is elicited by the experimenter. However, for enfacement illusion, there is a third possibility, that has never been studied, in which the participant moves the head actively by following the video, without an online pairing of the avatar and participants movements. In this way the participants do not have a real control over the avatar movements, but there is only an illusory control. Therefore, we decided to refer to this possibility in enfacement paradigm as guided movement. It is worth considering that even mere exposure to a body part can elicit embodiment (Dasgupta and Rivera, 2008; La Rocca et al., 2020). To our knowledge, the exposure condition has not yet been investigated with faces.

The literature reports different enfacement illusion paradigms. Some authors used two people sitting in front of each other (Sforza et al., 2010; Bufalari et al., 2019) others used movies displaying real unfamiliar faces (Tsakiris, 2008; Tajadura-Jiménez et al., 2012b; Panagiotopoulou et al., 2017; Gülbetekin et al., 2021) or humanoid animated characters (Gonzalez-Franco et al., 2020). More recently, the enfacement literature introduced the use of the 3-D personalized reconstruction of faces (Grewe et al., 2021) and other standardized avatars (Serino et al., 2015).

Using computer-generated (CG) faces has become increasingly common in different psychological research areas (Yaros et al., 2019). Artificial faces with a very human-like appearance can now be generated by several software programs (either "from scratch" or by inputting real photographs to be converted into 3-D head models). Once generated, the faces can then be manipulated for perceptual or psychological characteristics (e.g., expressions, viewpoint, emotions, and feature size).

Computer-generated faces can differ according to their human likeness, which describes the degree to which an entity has a human-like appearance and presents human physical traits. Furthermore, they can present different levels of photographic realism and physical appearance details (e.g., rendering, shades, and texture). The most notable difference between these CG faces and face photographs is that the CG faces appear to lack fine-grained surface texture information and imperfections that are usually present in photographic face stimuli. New recent software has been developed allowing the creation of highly realistic 3-D faces starting with face photographs (i.e., Character Creator).[1] To the best of our knowledge, no research in psychology has used this program to study face perception. Being able to edit the avatar's characteristics could influence the embodiment illusion experience (Tsakiris and Haggard, 2005; Lugrin et al., 2015) together with its perception and attitudes toward it (Peck et al., 2013).

The advantages of this type of stimuli is to expose participants to faces that are as realistic as possible and at the same time are editable under a variety of aspects (e.g., facial expressions, facial features, gender, and social cognition manipulation).

The present study aimed to (i) test the possibility of eliciting the enfacement illusion over virtual faces and (ii) verify which enfacement paradigm elicits a stronger illusion. Particularly, we aimed to verify whether simple exposure to faces without multisensory integration is sufficient in eliciting enfacement or if a multisensory and congruent stimulation is necessary. To do so, (i) Computer Generated faces were created through the software Character Creator and (ii) we compared enfacement and embodiment illusions in visuotactile stimulation, visuomotor stimulation, and exposure condition.

Having a methodological reference for studying enfacement through the use of avatars can lead to a wide range of applications in virtual reality experiments both in the cognitive and social domains.

2. Methods

2.1. Participants

A total of 24 young adults participants took part in the study. All participants were caucasian and we excluded those wearing

---

[1] https://www.reallusion.com/character-creator/

glasses and having beard, in order to have an homogenous group. The number of participants was calculated *a priori* through the software G*power 3.1.9.4. We referred to the recent literature about embodiment phenomena in VR (La Rocca et al., 2020; Tosi et al., 2020, 2021) that suggests a medium effect size for the experimental condition (eta-squared around 0.13). We run an *a priori* power analysis for a within-subjects repeated measure ANOVA encompassing a 2 × 3 × 2 design. The analysis revealed that to reach a power of 0.80, with alpha set to 0.05 and effect size set to 0.30, 24 participants were needed. The final number of participants is 23 [13 females, mean age = 25.47 (SD = 2.76), age range = 21–34], as the first participant was removed from the analyses due to a technical error. The reported research protocol was approved by the ethical committee of the University of Milano-Bicocca (protocol number: RM-2021-392), and written informed consent was obtained from all participants.

2.2. Stimuli

We generated four avatars, two males and two females, which had been shown, respectively to male and female participants. Avatars were created starting from two pictures belonging to the Chicago Face Database (Ma et al., 2015) having a suitability score above 4. Subsequently, they were morphed through the program Character Creator 3. Each photo underwent a digital transformation with the "Headshot" Plug-In. This plug-in can edit faces *via* pro-mode and auto-mode. The first one is designed for high-resolution texture processing and facial morph definition. Auto-mode makes a lower definition avatar but allows to generate 3-D hair starting from the original photo. We processed *via* auto-mode to generate 3-D hair and then converted our stimuli through pro-mode to obtain highly realistic faces.

Once the avatars were created, each one was inserted in an environment resembling the lab used for testing, edited through the "iClone 3DXChange 7" pipeline, and converted to.avi format videos. Videos were created to belong to one of the three experimental conditions: visuotactile stimulation, visuomotor stimulation, and simple exposure. Moreover, the respective control condition videos were created. **Figure 1** shows the avatars.

2.3. Procedure

Each participant underwent six conditions (experimental: synchronous visuotactile, synchronous visuomotor, synchronous exposure; control: asynchronous visuotactile, asynchronous visuomotor, asynchronous exposure). The experimental setting matched the lab environment re-created in the videos. Participants sat in front of a screen and, with the back leaned, the distance of the eyes from the monitor was approximately 50 cm. The screen height with respect to the floor was adjusted to resemble a mirror by aligning it with the participant's head.

The six conditions were administered in a counterbalanced order: we created 24 unique combinations of order presentation, and each one was presented to one of our participants. Each participant was exposed to one of the avatar matching his/her gender, whose identity was assigned in a counterbalanced order.

In the case of the synchronous visuotactile condition, the video showed an avatar that stayed still while a chopstick touched his/her cheek at a frequency of 1 Hz (the pace was given by the metronome) for 2 min. To ensure that the touch was realistic, the cheek was edited to reproduce the skin reaction to a touch in that position. While viewing the avatar being touched on his/her cheek by the chopstick, participants received a synchronous tactile stimulation by the experimenter. Touches were delivered on the corresponding location of the participant's cheek at a frequency of 1 Hz following the same pace given by the metronome in the video. In the asynchronous control visuotactile condition, the chopstick touched the avatar's cheek in random same positions and with an anti-phasic rhythm. During the video, the experimenter touched the participants' cheek with the same rhythm as in the synchronous condition. However, the effective touch did not match the video either with respect to the location or the rhythm of the observed touch. The metronome was still active to maintain equal circumstances.

In the case of the synchronous visuomotor condition, the video showed an avatar that was modified to produce either a nodding or a shaking guided movement. Half of the participants were presented with the nodding movement, the other half saw the shaking one. The movement was regular and followed the rhythm of a head movement per second. Participants were instructed to nod/shake their heads following the same pace given by the metronome in the video. In the asynchronous visuomotor condition, the avatar produced the movement (nodding or shaking) following a random rhythm. Participants received the same instructions as in the synchronous one, but the observed avatar did not match their movement (see Section "2.2. Stimuli"). Participants were instructed to move following the metronome. This served to create the illusion of controlling the avatar's movements in the synchronous condition. On the other hand, in the asynchronous condition it served to de-synchronize participants' and avatars' movements. However, there was not registration of participants' actual movements through face expressions and movements online trackers. The movements were externally guided.

In the congruent exposure condition, the avatar was presented as static and in the same position as the participants' faces. In the control incongruent exposure condition, the avatar was presented as static and inverted. During both the synchronous and control exposure conditions, participants were only instructed to look at the avatar for 2 min. Each video lasted 2 min.

In order to make the results comprehensible and comparable to the other condition, we will refer to the congruent and incongruent exposure conditions as synchronous and asynchronous.

After each condition, participants answered 16 self-report questions to assess their subjective experience during the video (a schematic representation of the procedure can be found in **Figure 2**). The first six questions (Q1–Q6) belonged to a questionnaire used for investigating the embodiment effect (Tosi et al., 2020, 2021; Tosi and Romano, 2022). Items were re-adapted to be specific for face stimuli. The following ten questions (Q7–Q16) belonged to the enfacement questionnaire (Tajadura-Jiménez et al., 2012b). We removed eight questions from the original enfacement questionnaire as they were specifically related to the visuotactile condition (i.e., questions 1 and 2) or to the visuomotor condition (i.e., question 8) or unrelatable according to

FIGURE 1
Males and females pictures of the used avatars.

our experimental paradigm (i.e., questions 11, 12, 15, 17, 18). The complete list of the questions is reported in **Table 1**.

Participants were asked to indicate their level of agreement with each question on a seven-point Likert scale (from−3—disagreement–to + 3—agreement). Before the analyses, we reversed the item Q4.

Participants were also administered the self-esteem IAT, as part of a wider project. However, the results will not be discussed in the present article.

The overall experimental design consisted of a 2 (Congruency) × 3 (Stimulation) × 2 (Questionnaire) within-subject design. The following dimensions were assessed: Congruency (i.e., synchronous experimental condition vs. asynchronous control condition), Stimulation (i.e., visuotactile, visuomotor, and exposition), and Questionnaire (embodiment items vs. enfacement items). The videos of the experimental conditions and the dataset are available on the Open Science Framework platform at the following link: https://osf.io/cf8qv/?view_only=efd1bb4b124a4c12b295c5f31ea8bc20.

2.4. Analyses

Before running the analyses, each participant's responses to the questionnaire have been ipsatized by centering the responses on the average score of all the questions in all the conditions and dividing the resulting value by the standard deviation of the whole set of responses. The procedure is a within-subject normalization and removes the response set bias (i.e., the participant's response style). Thus, each item is coded in terms of standard deviations from each participant's average response (Hofstede, 1984). We then clustered the first six items of the questionnaire (Q1–Q6) by averaging their values because they are all part of the main factor embodiment (Longo et al., 2008; Tajadura-Jiménez et al., 2012b; Romano et al., 2021; Tosi and Romano, 2022). We also clustered the remaining ten items of the questionnaire (Q7–Q16) by averaging their values because they belong to the original enfacement questionnaire (Tajadura-Jiménez et al., 2012b). To examine the subjective experience of embodiment elicited by the different stimulations (visuotactile vs. visuomotor vs. exposition), we ran a repeated measures analysis of variance (rmANOVA) with a within-subject design that covered a 2 (Congruency) × 3 (Stimulation) × 2 (Questionnaire) full-factorial model. The factor named Questionnaire controlled whether there were any differences between the embodiment and the enfacement constructs, as assessed by the respective items. Significant effects have been interpreted by inspecting 95% Confidence Intervals. The analyses investigating the different subcomponents of the embodiment sensation (i.e., Ownership–Q1-Q2; Agency–Q3-Q4; Location–Q5-Q6) are included in the **Supplementary material**. We ran a rmANOVA with a within-subject design that covered a 2 (Congruency) * 3 (Stimulation) model. We conducted two additional analyses to control for specific aspects of the experimental design. In the visuomotor condition, half of the participants saw the nodding movement, the other half saw the shaking one. To control for any influence of the

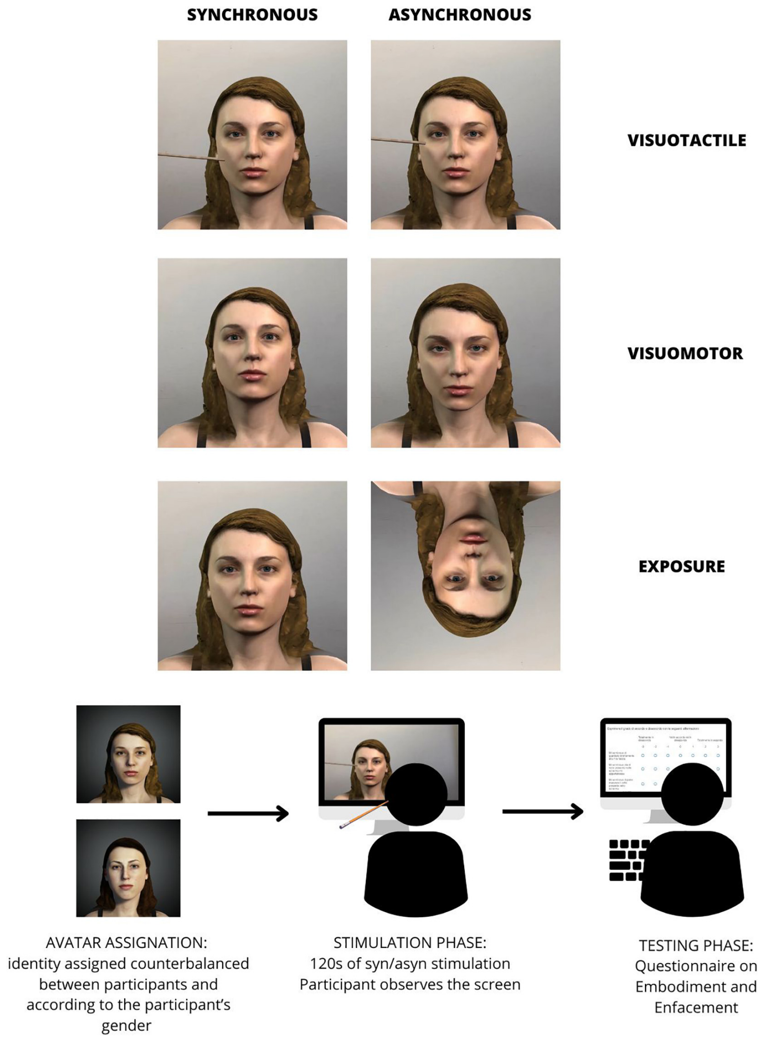


FIGURE 2
Description of the procedure of the experiment. Enfacement was elicited through three conditions, each in a synchronous (congruent) or asynchronous (incongruent) modality. Conditions were visuo-tactile stimulation, guided movement visuo-motor stimulation or simple exposure. Each participant was assigned one of two avatars in counterbalanced order to match his/her gender. After being exposed to each condition in each modality, participants completed the questionnaires about enfacement. Also stimulations were administered in counterbalanced order.

type of movement presented, we ran an rmANOVA with a mixed within-/between-subject design that covered a 2 (Congruency) × 2 (Questionnaire) × 2 (Type of movement) factorial model. The results of the analysis are reported in the **Supplementary materials**. As for the enfacement questionnaire, we specifically looked at item Q10 to assess whether the experimental design influenced the similarity participants perceived with the avatar. The results of the 2 (Congruency) * 3 (Stimulation) rmANOVA are reported in the **Supplementary materials**. We conducted all the analyses with the ezANOVA function for the statistical software R (R Core Team, 2017).

3. Results

We found significant main effects of Congruency [$F(1,22) = 82.28$, $p \leq 0.001$, $\eta^2_G = 0.40$] and Stimulation [$F(2,44) = 10.93$, $p = 0.001$, $\eta^2_G = 0.14$] (**Figure 3**). These results revealed greater embodiment values in the synchronous condition (CI: 0.16; 0.55) than in the asynchronous one (−0.53; −0.13). Moreover, participants showed higher embodiment sensation after the visuotactile (CI: −0.03; 0.39) and visuomotor (CI: −0.16; 0.32) stimulations as compared to the exposure condition (CI: −0.47; 0.03). We also found a significant interaction between Congruency and Questionnaire [$F(1,22) = 118.76$, $p \leq 0.001$, $\eta^2_G = 0.11$], suggesting that the embodiment statements caught a greater difference between the synchronous and asynchronous stimulations as compared to the enfacement items (**Figure 3**). Moreover, the interaction between Stimulation and Questionnaire resulted to be significant [$F(2,44) = 1.59$, $p \leq 0.05$, $\eta^2\ _G = 0.02$], showing greater embodiment ratings as compared to the enfacement one only after the visuotactile stimulation (**Figure 3**).

No further significant effects emerged (all other p-values > 0.15). The results obtained from each specific subcomponent of the embodiment construct

TABLE 1 Items of the embodiment and enfacement questionnaires.

Subcomponents		ID	Question
Embodiment	Ownership	Q1	It seems like I was looking directly at my own face
		Q2	It seems like the face in the video belonged to me
	Agency	Q3	It seems like I could have moved the face in the video?
		Q4	It seems like I was not in control of the face in the video?
	Location	Q5	It seems like the face in the video was in the location where my face was
		Q6	It seems like I could have felt a touch given to the face in the video
Enfacement		Q7	I felt like the other's face was my face
		Q8	It seemed like the other's face belonged to me
		Q9	It seed like I was looking at my own mirror reflection
		Q10	It seemed like the other's face began to resemble my own face
		Q11	It seemed like my own face began to resemble the other person's face
		Q12	It seemed like my own face was out of my control
		Q13	It seemed like the experience of my face was less vivid than normal
		Q14	It seemed like the person in the video was attractive
		Q15	It seemed like the person in the video was trustworthy
		Q16	I felt that the other person was imitating me

It might be noted that items are translated from Italian to English for publication purposes. However, Q5 might sounds unclear, but it was adapted referring to the position of the face and not its location in space.

and from the additional control analyses are reported in **Supplementary Tables 1–3**.

4. Discussion

The present study aimed to verify the possibility of eliciting enfacement over computer generated faces and test which stimulation condition elicits the stronger illusion. We created CG faces with the software Character Creator, a new generation software allowing the creation of high-quality texture 3-D objects and avatars starting from face photographs. Although other studies used avatars in embodiment paradigms, we aimed to systematically investigate different enfacement stimulation conditions on those types of stimuli. The ultimate goal was to set a methodological reference to apply enfacement to face processing research. The role of visuotactile stimulation in enfacement paradigms has already been extensively studied. However, a visuomotor stimulation, not requiring an acquisition system for online tracking, has been investigated less frequently. Moreover the role of mere exposure, to the best of our knowledge, has never been studied. In our study, the enfacement illusion was compared among visuotactile stimulation, visuomotor stimulation, and exposure condition. Each stimulation comprised a congruent (i.e., synchronous) and an incongruent (i.e., asynchronous) condition. Our results suggest a difference in congruency (synchronous vs. asynchronous) where congruent stimulation elicited higher enfacement effects than incongruent. Our results confirm that a multisensory stimulation, either visuotactile or visuomotor, administered with spatial and temporal congruency is able to elicit enfacement. As for the visuomotor stimulation, it is important to underline that we used a guided movement as visuomotor stimulation. As already stated, the advantage of active over non-active movement to elicit embodiment is not clear in the current literature (Dummer et al., 2009; Kalckert and Ehrsson, 2014). Our results confirm that even a guided movement as a visuomotor stimulation is able to elicit enfacement. The method we used has the advantage of being more accessible and feasible with respect to active visuomotor stimulation because it does not require any facial motion capture system or complete immersive virtual reality environment. This result is in line with previous studies about embodiment and enfacement illusions, where the synchronous condition elicits a stronger illusion effect as compared to the asynchronous one (Longo et al., 2008; Tsakiris, 2008; Kilteni et al., 2015; Porciello et al., 2018).

Crucially, our results indicate that even the mere exposure to a CG face elicits a stronger enfacement effect when the face is presented in a congruent position as compared with a reversed face.

Aside from investigating the role of congruency, we also directly assessed whether there was any difference between the different stimulations we used (i.e., visuotactile, visuomotor, and exposure). Results reveal that visuotactile and visuomotor stimulation conditions create a stronger illusion over the virtual face as compared to the simple exposure condition. Even if the mere exposure is enough to induce an embodiment effect, in line with La Rocca et al. (2020), the effect is significantly weaker as compared to multisensory stimulation.

Moreover, we observed a significant interaction between congruency and the used questionnaire. This result indicates that the adapted version of the embodiment questionnaire is more sensitive in capturing the difference between synchronous and asynchronous stimulation as compared to the enfacement questionnaire. Thus, it appears clear that methodological research on the enfacement questionnaires is still needed. For example, recent literature uses a self-recognition task on a continuum of morphed images ranging between two identities. This serves to investigate the level of enfacement with the seen avatars. In fact, this measure should implicitly tell us what is the identification of the participant with a different identity (Deltort et al., 2022). It would be interesting to use it in future studies as it is an implicit measure which could be best to avoid test-retest effects. This measure would be helpful to investigate enfacement also in clinical populations (Ferroni et al., 2019; Deltort et al., 2022).

A possible limitation of the present study is that we created for each gender two avatars that we assigned to participants. However, we did not control for the similarity of appearance of our avatars with the participants. As a matter of fact, Fribourg et al. (2020) find that avatar appearance impacts the sense of embodiment less than other dimensions, such as control over it and its point of view (Fribourg et al., 2020). The authors suggest that this result may

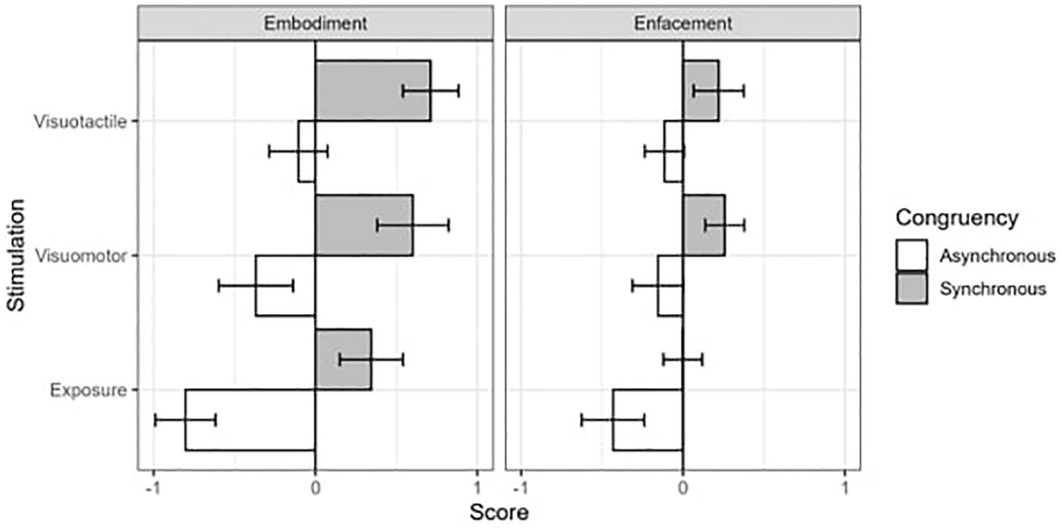


FIGURE 3
Results of the within-subjects 2 (Congruency) * 3 (Stimulation) * 2 (Questionnaire) repeated measure ANOVA on the averaged ipsatized answers to questionnaire statements. Gray and white columns display, respectively synchronous and asynchronous conditions. Error bars display confidence intervals.

depend on the task used. On the other hand, Waltemate et al. (2018) find that personalized avatars significantly increase body ownership and sense of presence.

To explore its potential role, we looked at Q10 from our questionnaire (i.e., “It seemed like the other’s face began to resemble my own face”). This result is described in **Supplementary Table 3** and suggests that in the synchronous condition, participants perceive the avatar as more similar to themselves than in the asynchronous condition. This result is interesting though only exploratory. Moreover, we cannot be sure whether the similarity between the participant and the avatar caused the embodiment effect or whether the perceived similarity was induced by the experimental manipulation. Thus, future research should quantitatively investigate the relationship between the similarity of the avatar to the participant and the enfacement effect. A further limitation of the present study regards the visuomotor stimulation condition. Two movements were used between participants: head nodding and shaking. We decided to use them both to avoid our results being driven by the potential valence of movements. In a control analysis, we checked whether there was a difference in enfacement scores depending on the presented movement. We expected the movements to equally elicit enfacement: however, we found higher scores in the group presented with the shaking movement. Nevertheless, this difference did not alter the main results of the present study. This analysis is reported in **Supplementary Table 2**. This result could be due to a difference in the foveal representation of the observed faces during stimulation, with the shaking face being easier to keep under fixation. However, this hypothesis is only speculative as we do not have enough data to drive conclusions. This result suggests that future studies should be careful in choosing the specific movement for visuomotor stimulation.

In conclusion, computer-generated faces can be a valid alternative to real faces to elicit enfacement. Moreover, their suitability is proved even for visuotactile and visuomotor stimulation conditions. From a procedure point of view, the novelty of the study is that it verifies enfacement illusion even in a setting of augmented reality. This makes it possible for other researchers not to use a VR headset or a completely immersive procedure but just a computer screen. Being able to embody a face of an avatar opens a wide range of possibilities in face processing research. In fact, the software used in the present experiment allows the manipulation of characteristics of the avatar’s face in virtual reality environments such as facial features appearance (i.e., different configurations of features and sizes), facial expressions, ethnicity, gender spectrum, and age span.

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://osf.io/cf8qv/?view_only=efd1bb4b124a4c12b295c5f31ea8bc20.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethical Committee of the University of Milano–Bicocca protocol number: RM-2021-392. The patients/participants provided their written informed consent to participate in this study.

Author contributions

SL and SG provided the initial conception, organization, and main writing of the text. GT analyzed the data, prepared all figures and tables, and contributed to the writing of the text. EF collected the data. RD supervised the study and read and approved the draft. All authors contributed to the article and approved the submitted version.

Acknowledgments

The authors thank the participants, who kindly participated in this research.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2023.1026196/full#supplementary-material

References

Alzueta, E., Melcón, M., Jensen, O., and Capilla, A. (2020). The 'Narcissus Effect': Top-down alpha-beta band modulation of face-related brain areas during self-face processing. *NeuroImage* 213:116754. doi: 10.1016/j.neuroimage.2020.116754

Apps, M. A., Tajadura-Jiménez, A., Sereno, M., Blanke, O., and Tsakiris, M. (2015). Plasticity in unimodal and multimodal brain areas reflects multisensory changes in self-face identification. *Cereb. Cortex* 25, 46–55. doi: 10.1093/cercor/bht199

Bortolon, C., and Raffard, S. (2018). Self-face advantage over familiar and unfamiliar faces: A three-level meta-analytic approach. *Psychon. Bull. Rev.* 25, 1287–1300. doi: 10.3758/s13423-018-1487-9

Bufalari, I., Sforza, A. L., Di Russo, F., Mannetti, L., and Aglioti, S. M. (2019). Malleability of the self: Electrophysiological correlates of the enfacement illusion. *Sci. Rep.* 9:1682. doi: 10.1038/s41598-018-38213-y

Dasgupta, N., and Rivera, L. M. (2008). When social context matters: The influence of long–term contact and short–term exposure to admired outgroup members on implicit attitudes and behavioral intentions. *Soc. Cogn.* 26, 112–123. doi: 10.1521/soco.2008.26.1.112

Deltort, N., Swendsen, J., Bouvard, M., Cazalets, J.-R., and Amestoy, A. (2022). The enfacement illusion in autism spectrum disorder: How interpersonal multisensory stimulation influences facial recognition of the self. *Front. Psychiatry* 13:946066. doi: 10.3389/fpsyt.2022.946066

Devue, C., and Brédart, S. (2011). The neural correlates of visual self-recognition. *Conscious. Cogn.* 20, 40–51. doi: 10.1016/j.concog.2010.09.007

Dummer, T., Picot-Annand, A., Neal, T., and Moore, C. (2009). Movement and the rubber hand illusion. *Perception* 38, 271–280. doi: 10.1068/p5921

Ferroni, F., Ardizzi, M., Sestito, M., Lucarini, V., Daniel, B., Paraboschi, F., et al. (2019). Shared multisensory experience affects Others' boundary: The enfacement illusion in schizophrenia. *Schizophr. Res.* 206, 225–235. doi: 10.1016/j.schres.2018.11.018

Fribourg, R., Argelaguet, F., Lécuyer, A., and Hoyet, L. (2020). Avatar and sense of embodiment: Studying the relative preference between appearance, control and point of view. *IEEE Trans. Vis. Comput. Graph.* 26, 2062–2072. doi: 10.1109/TVCG.2020.2973077

Gonzalez-Franco, M., Steed, A., Hoogendyk, S., and Ofek, E. (2020). Using facial animation to increase the enfacement illusion and avatar self-identification. *IEEE Trans. Vis. Comput. Graph.* 26, 2023–2029. doi: 10.1109/TVCG.2020.2973075

Grewe, C. M., Liu, T., Kahl, C., Hildebrandt, A., and Zachow, S. (2021). Statistical learning of facial expressions improves realism of animated avatar faces. *Front. Virtual Reality* 2:619811. doi: 10.3389/frvir.2021.619811

Gülbetekin, E., Altun, E., Er, M. N., Fidancı, A., Keskin, P., and Steenken, D. (2021). Effects of right or left face stimulation on self and other perception in enfacement illusion. *Soc. Neurosci.* 16, 189–205. doi: 10.1080/17470919.2021.1886983

Hofstede, G. (1984). *Culture's consequences: international differences in work-related values*, Vol. 5. Beverly Hills, CA: SAGE.

Kalckert, A., and Ehrsson, H. H. (2014). The moving rubber hand illusion revisited: Comparing movements and visuotactile stimulation to induce illusory ownership. *Conscious. Cogn.* 26, 117–132. doi: 10.1016/j.concog.2014.02.003

Kilteni, K., Maselli, A., Kording, K. P., and Slater, M. (2015). Over my fake body: Body ownership illusions for studying the multisensory basis of own-body perception. *Front. Hum. Neurosci.* 9:141. doi: 10.3389/fnhum.2015.00141

La Rocca, S., Brighenti, A., Tosi, G., and Daini, R. (2020). No country for old men: Reducing age bias through virtual reality embodiment. *Annu. Rev. Cyberther. Telemed.* 18, 127–131.

Lenggenhager, B., Tadi, T., Metzinger, T., and Blanke, O. (2007). Video ergo sum: Manipulating bodily self-consciousness. *Science* 317, 1096–1099. doi: 10.1126/science.1143439

Longo, M. R., Schüür, F., Kammers, M. P., Tsakiris, M., and Haggard, P. (2008). What is embodiment? A psychometric approach. *Cognition* 107, 978–998. doi: 10.1016/j.cognition.2007.12.004

Lugrin, J.-L., Latt, J., and Latoschik, M. E. (2015). "Avatar anthropomorphism and illusion of body ownership in VR," in *Proceedings of the 2015 IEEE virtual reality (VR)*, (Piscataway, NJ: IEEE), 229–230. doi: 10.1109/VR.2015.7223379

Ma, D. S., Correll, J., and Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behav. Res. Methods* 47, 1122–1135. doi: 10.3758/s13428-014-0532-5

Ma, Y., and Han, S. (2010). Why we respond faster to the self than to others? An implicit positive association theory of self-advantage during implicit face recognition. *J. Exp. Psychol. Hum. Percept. Perform.* 36:619. doi: 10.1037/a0015797

Malaspina, M., Albonico, A., Lao, J., Caldara, R., and Daini, R. (2018). Mapping self-face recognition strategies in congenital prosopagnosia. *Neuropsychology* 32:123. doi: 10.1037/neu0000414

Panagiotopoulou, E., Filippetti, M. L., Tsakiris, M., and Fotopoulou, A. (2017). Affective touch enhances self-face recognition during multisensory integration. *Sci. Rep.* 7:12883. doi: 10.1038/s41598-017-13345-9

Peck, T. C., Seinfeld, S., Aglioti, S. M., and Slater, M. (2013). Putting yourself in the skin of a black avatar reduces implicit racial bias. *Conscious. Cogn.* 22, 779–787. doi: 10.1016/j.concog.2013.04.016

Porciello, G., Bufalari, I., Minio-Paluello, I., Di Pace, E., and Aglioti, S. M. (2018). The 'Enfacement'illusion: A window on the plasticity of the self. *Cortex* 104, 261–275. doi: 10.1016/j.cortex.2018.01.007

R Core Team (2017). *R: A language and environment for statistical computing [Computer software] (Version 3.5. 1).* Vienna: R Foundation for Statistical Computing.

Romano, D., Maravita, A., and Perugini, M. (2021). Psychometric properties of the embodiment scale for the rubber hand illusion and its relation with individual differences. *Sci. Rep.* 11:5029. doi: 10.1038/s41598-021-84595-x

Serino, A., Sforza, A. L., Kanayama, N., van Elk, M., Kaliuzhna, M., Herbelin, B., et al. (2015). Tuning of temporo-occipital activity by frontal oscillations during virtual mirror exposure causes erroneous self-recognition. *Eur. J. Neurosci.* 42, 2515–2526. doi: 10.1111/ejn.13029

Sforza, A., Bufalari, I., Haggard, P., and Aglioti, S. M. (2010). My face in yours: Visuo-tactile facial stimulation influences sense of identity. *Soc. Neurosci.* 5, 148–162. doi: 10.1080/17470910903205503

Sugiura, M., Watanabe, J., Maeda, Y., Matsue, Y., Fukuda, H., and Kawashima, R. (2005). Cortical mechanisms of visual self-recognition. *Neuroimage* 24, 143–149. doi: 10.1016/j.neuroimage.2004.07.063

Tajadura-Jiménez, A., Grehl, S., and Tsakiris, M. (2012a). The other in me: Interpersonal multisensory stimulation changes the mental representation of the self. *PLoS One* 7:e40682. doi: 10.1371/journal.pone.0040682

Tajadura-Jiménez, A., Longo, M. R., Coleman, R., and Tsakiris, M. (2012b). The person in the mirror: Using the enfacement illusion to investigate the experiential structure of self-identification. *Conscious. Cogn.* 21, 1725–1738. doi: 10.1016/j.concog.2012.10.004

Tajadura-Jiménez, A., Lorusso, L., and Tsakiris, M. (2013). Active and passive-touch during interpersonal multisensory stimulation change self–other boundaries. *Conscious. Cogn.* 22, 1352–1360. doi: 10.1016/j.concog.2013.09.002

Tosi, G., and Romano, D. (2022). The network of the subjective experience in embodiment phenomena. *Psychol. Res.* doi: 10.1007/s00426-022-01714-7

Tosi, G., Maravita, A., and Romano, D. (2021). I am the metre: The representation of one's body size affects the perception of tactile distances on the body. *Q. J. Exp. Psychol.* 75, 583–597. doi: 10.1177/17470218211044488

Tosi, G., Parmar, J., Dhillon, I., Maravita, A., and Iaria, G. (2020). Body illusion and affordances: The influence of body representation on a walking imagery task in virtual reality. *Exp. Brain Res.* 238, 2125–2136. doi: 10.1007/s00221-020-05874-z

Tsakiris, M. (2008). Looking for myself: Current multisensory input alters self-face recognition. *PLoS One* 3:e4040. doi: 10.1371/journal.pone.0004040

Tsakiris, M. (2010). My body in the brain: A neurocognitive model of body-ownership. *Neuropsychologia* 48, 703–712. doi: 10.1016/j.neuropsychologia.2009.09.034

Tsakiris, M., and Haggard, P. (2005). The rubber hand illusion revisited: Visuotactile integration and self-attribution. *J. Exp. Psychol. Hum. Percept. Perform.* 31:80. doi: 10.1037/0096-1523.31.1.80

Waltemate, T., Gall, D., Roth, D., Botsch, M., and Latoschik, M. E. (2018). The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE Trans. Vis. Comput. Graph.* 24, 1643–1652. doi: 10.1109/TVCG.2018.2794629

Yaros, J. L., Salama, D. A., Delisle, D., Larson, M. S., Miranda, B. A., and Yassa, M. A. (2019). A memory computational basis for the other-race effect. *Sci. Rep.* 9:19399. doi: 10.1038/s41598-019-55350-0

frontiers | Frontiers in Human Neuroscience

TYPE Review
PUBLISHED 21 April 2023
DOI 10.3389/fnhum.2023.1077409

OPEN ACCESS

EDITED BY
Elena Nava,
University of Milano-Bicocca, Italy

REVIEWED BY
Navpreet Mann,
Shree Guru Gobind Singh Tricentenary University, India
Didier A. Depireux,
University of Maryland, College Park, United States

*CORRESPONDENCE
Karolina Kluk
✉ karolina.kluk@manchester.ac.uk

RECEIVED 22 October 2022
ACCEPTED 30 March 2023
PUBLISHED 21 April 2023

CITATION
Aloufi N, Heinrich A, Marshall K and Kluk K (2023) Sex differences and the effect of female sex hormones on auditory function: a systematic review.
Front. Hum. Neurosci. 17:1077409.
doi: 10.3389/fnhum.2023.1077409

Sex differences and the effect of female sex hormones on auditory function: a systematic review

Nada Aloufi[1,2], Antje Heinrich[1], Kay Marshall[3] and Karolina Kluk[1]*

[1]Manchester Centre for Audiology and Deafness, Manchester Academic Health Science Centre, Faculty of Biology, Medicine and Health, School of Health Sciences, University of Manchester, Manchester, United Kingdom, [2]College of Medical Rehabilitation Sciences, Taibah University, Medina, Saudi Arabia, [3]Division of Pharmacy and Optometry, Faculty of Biology, School of Health Sciences, Medicine and Health, University of Manchester, Manchester, United Kingdom

Aims: First, to discuss sex differences in auditory function between women and men, and whether cyclic fluctuations in levels of female sex hormones (i.e., estradiol and progesterone) affect auditory function in pre-menopausal and post-menopausal women. Second, to systematically review the literature concerning the discussed patterns in order to give an overview of the methodologies used in research. Last, to identify the gap in knowledge and to make recommendations for future work.

Methods for the systematic review: Population, Exposure, Control, Outcome and Study design (PECOS) criteria were used in developing the review questions. The review protocol follows the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) and was pre-registered in the Prospective Register of Systematic Reviews (PROSPERO; CRD42020201480). Data Sources: EMBASE, PubMed, MEDLINE (Ovid), PsycINFO, ComDisDome, CINAHL, Web of Science, Cochrane Central Register of Controlled Trials (CENTRAL) via Cochrane Library, and scanning reference lists of relevant studies, and internet resources (i.e., Mendeley) were used. Only studies published between 1999 and 2022, in English, or in English translation, were included. The quality of evidence was assessed using the Newcastle-Ottawa Scale (NOS).

Results: Sex differences: Women had more sensitive hearing (measured at the level of peripheral and central auditory system) than men. Cyclic fluctuations: Auditory function in women fluctuated during the menstrual cycle, while no such fluctuations in men over the same time period were reported. Hearing sensitivity improved in women during the late follicular phase, and decrease during the luteal phase, implying an effect of female sex hormones, although the specific effects of estradiol and progesterone fluctuations on the central auditory system remain unclear. Hearing sensitivity in women declined rapidly at the onset of menopause.

Conclusion: The review has shown the following. Consistent sex differences exist in auditory function across the auditory pathway with pre-menopausal women often showing better function than age-matched men. Moreover, pre-menopausal women show fluctuations in hearing function across the menstrual cycle with a better function during the peak of estradiol or when the ratio of estradiol to progesterone is high. Third, menopause marks the onset of hearing loss in women, characterized by a rapid decline in hearing sensitivity and a more pronounced loss than in age-matched men. Finally, the systematic review highlights the need for well-designed and -controlled studies to evaluate the influence of estradiol and progesterone on hearing by consistently including control groups (e.g., age-matched man), using objective tests to measure hormonal levels (e.g., in saliva or blood), and by testing participants at different points across the menstrual cycle.

Systematic review registration: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020201480, identifier CRD42020201480.

KEYWORDS

sex differences, female hormones, auditory function, fluctuation, progesterone, estradiol

Definitions

Pre-menopause	The reproductive period of a woman's life.
Menopause	A time period of 12 consecutive months after the cessation of a woman's menstrual cycle.
Post-menopause	The time period after menopause, extended time period of amenorrhea.
Amenorrhea	The absence of menstruation.
Menstrual cycle phases	Phases of menstrual cycle are defined in this review relative to a typical 28-day cycle (in shorter cycles the follicular phase is attenuated and elongated in longer cycles) and are described as: •*Early Follicular phase:* day 1–8 of the cycle, where day 1 is the first day of menses (start of menstrual cycle) •*Late Follicular phase:* day 9–16 of the cycle (day 14 – ovulation) •*Early Luteal phase:* day 17–22 of the cycle •*Late Luteal phase:* day 23–28 of the cycle
Pure Tone Audiometry (PTA)	Behavioral test used to assess hearing sensitivity. Typically measured at 0.25, 0.5, 1, 2, 4, 6, and 8 kHz.
Oto-Acoustic Emissions test (OAEs)	Physiological test used to assess the health of outer hair cells (OHCs) in the cochlea by recording soft sounds emitted by the ear. OAEs can be spontaneous, i.e., Spontaneous Oto-Acoustic Emissions (SOAEs), and evoked by a click/tone burst (Transient Evoked Oto-Acoustic Emissions, TEOAESs), or by a combination of two tones (Distortion Product Oto-Acoustic Emissions (DPOAEs).
Auditory Brainstem Response (ABR)	Physiological measure of auditory pathway's neuroelectric activity from the auditory nerve to the cerebral cortex. ABRs can be evoked by a range of stimuli such as clicks, tone-bursts or complex stimuli.
Speech – ABR	ABR evoked by speech (e.g.,/ba/). Speech-ABR provides critical information on how more complex stimuli are processed by the brainstem.
Speech audiometry	Behavioral test used to assess speech recognition threshold (SRT), i.e., at what sound level does the speech need to be presented to be accurately perceived in fifty percent of the cases; and word recognition score (WRS), i.e., what proportion of words is accurately perceived at a particular presentation level of sound. Additional to speech perception, speech discrimination and comprehension can be tested to assess the ability to discriminate between similar words and comprehend sentences and continuous speech.
Women	Women are defined in this review as an adult who was identified as female at birth. This was chosen as most of the literature reviewed was published before the definition of the word "women" in Cambridge Dictionary was expanded (October 2022) to include an adult who lives and identifies as female though they may have been identified with a different sex at birth. Consequently, in this review the words "female" and "women" have the same definitions.
Men	Men are defined in this review as an adult who was identified as male at birth. This was chosen as most of the literature reviewed was published before the definition of the word "men" in Cambridge Dictionary was expanded (October 2022) to include an adult who lives and identifies as a male though they may have been identified with a different sex at birth. Consequently, in this review the words "male" and "men" have the same definitions.

1. Introduction

This is a narrative review followed by a systematic review of the available evidence on sex differences in auditory function, and the effect of changes in female sex hormone levels on hearing. By identifying these sex differences, researchers and clinicians will be able to understand the static impact of sex on different audiometric measures, and impact of dynamic fluctuations in sex hormones on hearing function. In addition, this review highlights the methodological concerns in research studies investigating sex differences and/or the effect of sex hormones on hearing. This can be used to improve future work in this field. Lastly, this review highlights the questions for which the available evidence provides a clear possibility of hormonal treatment for preserving hearing sensitivity in older women.

Sex differences in hearing have been reported by some (McFadden, 1993; Stuart and Kerls, 2018; Zakaria et al., 2019) but not others (Wadnerkar et al., 2008; Boothalingam et al., 2018). It is not clear whether these differences are genuine and occur due to biological sex differences (such as differences in sex hormones) or whether they are due to systematic differences between the sexes in exposure to environmental noise and/or ototoxins. Three pieces of evidence support the hypothesis that these differences are due to biological differences and that female sex hormones contribute to sex differences in hearing. Firstly, Turner syndrome patients (young women with abnormally low levels of female sex hormones) present with hearing thresholds comparable to those of women in control population at least 20 years older than their age group (Bonnard et al., 2017, 2018), which points to a protective role of sex hormones. Secondly, better hearing sensitivity in young women compared to age-matched men (McFadden et al., 2006) disappears, i.e., hearing sensitivity in women decreases, when women reach menopause (reduction in female sex hormones). Thirdly, women's hearing function fluctuates cyclically in synchrony with fluctuations in female sex hormones (Al-Mana et al., 2010). Moreover, biological sex has been reported to be associated with some aspects of cochlear function and its vulnerability to changes due to age or noise exposure. This is in particular due to the protecting mechanism of female sex hormones against noise exposure, and delaying the onset of age-related hearing loss in women (McFadden, 1998; Zündorf et al., 2011; Grinn et al., 2017; Shuster et al., 2019). In the following sections we will review first sex hormones in women and men, followed by description of auditory anatomy, functioning of the relevant sex hormones, and finish with a discussion of evidence for and against static and dynamic differences in hearing associated with sex hormones. We will then systematically review the literature to provide an overview of the methods and outcome measures used in the field. Finally we will summarize gaps of knowledge in the field and suggest potential ways forward.

1.1. Sex hormones

Hormone status differs between women and men during the reproductive years of life, both in the overall levels of hormones and in terms of regular fluctuations over time. In general, similar sex hormones (i.e., estradiol, progesterone, and testosterone) can be found both in women and men, however, the production sites, blood concertation, and their effect on different organs and systems differ greatly (Svechnikov and Söder, 2007). In women, estradiol (the most potent of the three naturally occurring estrogens) and progesterone are secreted by the ovaries in a cyclic pattern of high/low amounts (across the reproductive cycle), while testosterone is produced only in small amounts by both ovaries and the adrenal glands (Svechnikov and Söder, 2007). In men, high amounts of testosterone are secreted by the testes, while small amounts of estradiol and progesterone are produced by both the testes and adrenal glands (Tyagi et al., 2017). In men, hormone levels are relatively stable (Lauretta et al., 2018), while in women hormone levels fluctuate across the reproductive cycle and change across the lifespan. Estradiol is made in the adrenal glands, ovaries, and fat cells, and is found in both sexes, but its concentration in blood is higher in women than men. While the levels of estradiol fluctuate during the different stages of a woman's life (i.e., menstrual cycle, during pregnancy, and menopause), in men the level of this hormone remains largely stable (Lauretta et al., 2018). Progesterone, which is produced by the corpus luteum (He and Ren, 2018), counters the function of estradiol in non-pregnant women. It is mainly responsible for stimulating the ovaries to develop a new menstrual cycle and preparing the endometrium for implantation of the fertilized egg, thus its levels rise in the luteal phase (Simonoska et al., 2009). Progesterone is also the dominant hormone during pregnancy, as the placenta takes over the function of corpus luteum to secrete progesterone (He and Ren, 2018). Prolonged changes in hormone status, for instance during pregnancy when progesterone dominates, or menopause when overall sex hormone levels decline, have been associated with reduced hearing sensitivity (e.g., Guimaraes et al., 2006; Al-Mana et al., 2008; Emami et al., 2018).

The reproductive time span in women begins at menarche (pre-menopausal) and ends when the menstrual cycle ceases (amenorrhea) and a woman enters a period called menopause. When amenorrhea lasts for longer than 12 consecutive months, a woman enters a period called post-menopause. There are distinct hormonal changes that coincide with these three phases of a non-pregnant woman's reproductive cycle: In pre-menopausal period, the amount of female sex hormones (e.g., estradiol and progesterone) fluctuates cyclically during the menstrual cycle.

The following section outlines the cyclical characteristics of the ovarian cycle and is followed by a discussion of the two hormones with particular relevance to hearing: estradiol and progesterone.

1.1.1. The ovarian cycle

Hormonal regulation in both women and men is controlled by the hypothalamus and the pituitary gland. The main difference in hormonal regulation between the sexes is the frequency of change, i.e., women go through a full female reproductive cycle each month in addition to changes that occur across the lifespan, while in men hormonal changes occur only across the lifespan. The hypothalamus in the female brain produces the gonadotropin-releasing hormone (GnRH), which causes the anterior pituitary gland to produce two hormones (gonadotrophins) that are essential to the ovarian cycle: follicle stimulating hormone (FSH) and luteinizing hormone (LH) (Hawkins and Matzuk, 2008). The concentration of these hormones fluctuates during the menstrual cycle as estradiol has a feedback action upon their release. In most of the cycle, estradiol exerts homeostatic negative feedback on GnRH (Moenter et al., 2009).

The average length of the cycle is 28 days (Najmabadi et al., 2020), and it can be divided into four phases: early follicular phase defined as day 1–8 of the cycle, where day 1 is the first day of menses (start of menstrual cycle); late follicular phase defined as day 9–16 of the cycle (day 14 – ovulation); early luteal phase defined as day 17–22 of the cycle; and late luteal phase defined as day 23–28 of the cycle (see **Figure 1**). Ovulation typically finishes by Day 16 (i.e., 12–14 days before the next menstrual cycle begins). These phases are dominated by different hormones, two of which are of particular interest to hearing: estradiol and progesterone.

In the beginning of the menstrual cycle (early follicular phase), the concentration of estradiol is low. This low concentration of estradiol inhibits the secretion of LH and slightly increases the release of FSH (see **Figure 1**).

At the end of the early follicular phase, estradiol levels rise leading to positive feedback and the release of GnRH, which in turn activates LH and FSH to surge and initiate ovulation. The pulsatile nature of the release of GnRH determines the ratio of release of the two gonadotrophins. The level of estradiol fluctuates during the menstrual cycle and reaches its peak in the late follicular phase when FSH enters the ovary and helps the primary follicle to develop into a secondary follicle. High levels of estradiol cause positive feedback and the release of LH, which consequently increases its secretion. The high level of LH triggers ovulation and the release of the mature follicle (usually at day 14 in a 28-day cycle).

After ovulation (day 14), the early luteal phase begins in which the level of LH drops dramatically as the level of estradiol decreases (Hawkins and Matzuk, 2008). The corpus luteum (i.e., the remains of the follicle) produces estradiol and progesterone, and progesterone level starts to increase in this phase until it reaches its peak around day 21 (Hawkins and Matzuk, 2008; see **Figure 1**).

During the early and late luteal phases, progesterone plays an important role in inhibiting the secretion of GnRH in the hypothalamus, in order to prevent the release of FSH and LH and stop the development of a new cycle. Therefore, the levels of GnRH, FSH and LH decrease in the early and late luteal phases (Hawkins and Matzuk, 2008).

The levels of these sex hormones can be assessed using different methods. The most accurate measures are biological measures such as blood or saliva samples where hormone levels are directly assessed. Alternatively, self-report measures can be used, however, this might not be an accurate or consistent measure (Farrar et al., 2015).

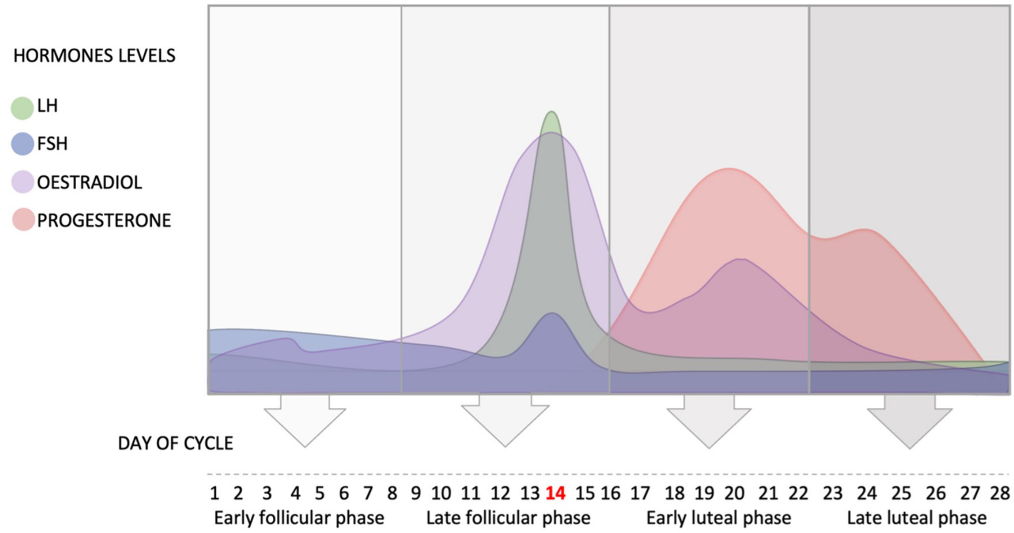


FIGURE 1
Schematic representation of the fluctuation of the hypothalamus and ovarian hormones during the average ovarian cycle, and the four phases of the cycle. FSH, follicle stimulating hormone; LH, luteinizing hormone.

1.2. Auditory pathway

The human auditory system comprises the peripheral auditory pathways: the external, middle, inner ears, and the vestibulocochlear nerve (8th cranial nerve), which connects to the central nervous system (McFadden, 1998; Zündorf et al., 2011; Grinn et al., 2017; Shuster et al., 2019) and central auditory pathways: cochlear nuclei, superior olivary nuclei, lateral lemniscus, inferior colliculus, medial geniculate nuclei, and auditory cortex (Shuster et al., 2019). Function of the auditory system can be assessed using either behavioral or physiological measures.

1.2.1. Behavioral measures of the auditory function

Behavioral methods require participants' active cooperation to provide responses. For example, Pure Tone Audiometry (PTA) is a behavioral hearing sensitivity measure that requires participants to indicate (by pressing a response button) when they heard the test sound. Speech Audiometry tests require participants to repeat speech samples that they heard. They provide information on how well the auditory system processes speech signals, which are more similar to natural signals heard in daily life than the pure tones used in PTA. Speech audiometry can be carried out for speech presented in quiet or in background noise. The latter can be useful for assessing not only auditory function but also cortical speech and language function.

1.2.2. Physiological measures of auditory function

Physiological responses are recorded without the need for participants' active cooperation. For example, Otoacoustic Emissions (OAEs), a marker of the health of the outer hair cells (OHCs) in the cochlea, are recorded from the participants' ears without any need for participants' active cooperation (Gold, 1948; Kemp, 1978; Gelfand, 2004; Grabham et al., 2013). Different types of OAEs can be recorded that reflect slightly different aspects of function of OHCs in the cochlea (e.g., Robinette and Glattke, 1997; Grabham et al., 2013). Spontaneous Otoacoustic Emission (SOAE) is a constant unprompted sound emitted from the cochlea that is always present, without any external stimulus. The sound pressure levels of SOAE range between 10 and 30 dBSPL, i.e., they are not usually audible to those who have them (Kemp, 2002). SOAEs are one sign of a healthy cochlea, are spontaneously produced, and are present in 30% (Robinette and Glattke, 1997) to 70% (Abdala and Visser-Dumont, 2001) of all listeners with normal hearing. In contrast, Transient Evoked Otoacoustic Emissions (TEOAEs) are evoked OAEs and can therefore be easily elicited from all healthy ears. TEOAEs are evoked by a short click stimulus and emit complex signals back to the external auditory meatus milliseconds after its presentation. Distortion Product Otoacoustic Emissions (DPOAE) are a third type of OAEs that can be used in order to assess the health of the cochlea. They are evoked when two tones of different frequencies (f_1 and f_2) are presented to the ear and the ear emits back distortion products of the presented sounds. The DPOAE that can be detected most prominently occurs at a frequency equal to $2f_1$–f_2 of the presented sound.

Auditory Brainstem Responses (ABRs) are electrophysiological measures of hearing sensitivity and auditory function up to the level of brainstem, do not require patient's cooperation, and are often measured while the patient is asleep (Corwin et al., 1982). ABRs consist of five waves, with each wave originating from a different part of the central auditory system, starting with the spiral ganglion in the cochlea (wave I) all the way to the inferior colliculus (wave V) (McFadden, 1998).

1.3. Anatomical evidence for the influence of sex hormones on hearing

1.3.1. Estradiol (E2, or 17β-estradiol) receptors

The following section will discuss anatomical evidence showing that hearing function can be affected by sex hormones. As a first step it is important to note that estradiol has impact beyond the reproductive system and influences the physiological function of other body organs and systems such as the skeletal, cardiovascular, and nervous systems (Al-Mana et al., 2008; He and Ren, 2018). In hearing, estradiol may improve the inflow of metabolites to inner ear cells, which is vital for processing of auditory signals, and has been found to act as a neuromodulator in facilitating detection of auditory signals (Tremere et al., 2009).

Estradiol receptors (ERs) have been found in the inner ear of both animals and humans. The role of ERs is to mediate the effect of estradiol in the cells. Two types of intracellular estradiol receptor exist, ERα and ERβ. ERα are likely to influence the cochlear and vestibular sensory transduction, while ERβ may have more central, neuroprotective role (Meltser et al., 2008).

In human studies, ERα have only been found in the spiral ganglion, and ERβ only in stria vascularis cells, which are essential to signal transmission and cochlear homeostasis, respectively, (Stenberg et al., 2001). The presence of ERα and ERβ in the ear affect auditory function in humans in a number of ways. First, estradiol receptors mediate the role of estradiol on the neuronal plasticity, and the metabolic levels of neurotransmitters and blood flow (Stenberg et al., 1999; Caruso et al., 2000; Lee and Marcus, 2001). Second, while ERα and ERβ are found in both men and women, their expression is related to the level of estradiol in the blood serum (Hultcrantz et al., 2006; Motohashi et al., 2010), and this level fluctuates over time in women. Additionally, the up- and down-regulation of ERα and ERβ in the inner ear depends on the life stage of a women (Simonoska et al., 2009), such that the level of estradiol influences auditory function in different ways at different times, in particular during maturation of the organism, the menstrual cycle, pregnancy, and menopause (Al-Mana et al., 2008; He and Ren, 2018).

In animals, ERα and ERβ have been found in the inner ear plasma membrane cells, the cochlear and vestibular fluids, cochlear cells including the OHCs, inner hair cells (IHCs), stria vascularis, spiral ligament, Reissner's membrane, and spiral ganglion cells (Stenberg et al., 1999), and distributed throughout the whole auditory pathway (Stenberg et al., 1999; Charitidi et al., 2009, 2010; Charitidi and Canlon, 2010). Estradiol receptors have also been found in the central nervous system (Contoreggi et al., 2021). In mice, ERα and ERβ were found in the ventral cochlear nucleus, nucleus of the trapezoid body, the lateral- and medio-ventral periolivary nuclei, the dorsal lateral lemniscus, and the inferior colliculus. In lateral olive, the ventral lateral lemniscus and central nucleus of the inferior colliculus only ERβ were found and in the auditory cortex only ERα were found (Charitidi et al., 2010). Similar to human studies, animals showed better hearing sensitivity during higher levels of estradiol (e.g., Sisneros et al., 2004; Arch and Narins, 2009; Frisina, 2012). No sex differences have been found in the expression patterns of estradiol receptors in the central auditory system neither in young nor aged mice (Charitidi and Canlon, 2010; Charitidi et al., 2010).

Regarding potential underlying mechanisms, estradiol has been suggested to play a role in aiding neural excitation in the inner ear and increase the neurosteroids in the brainstem, enhancing the transmission of the auditory signals to the brain (Tremere et al., 2011).

1.3.2. Progesterone receptors

In contrast to estradiol, there is no evidence of the presence of progesterone receptors in the inner ear in either humans or rats (Bonnard et al., 2013). No staining of progesterone receptors observed in stria vascularis, the organ of Corti or the spiral ganglion in either human or rat inner ears. However, progesterone receptor-B was found in the cochlear bone (Bonnard et al., 2013).

1.4. Functional evidence for the influence of sex hormones on hearing

The human auditory system shows a number of minor but significant functional sex differences (McFadden, 1998). These differences can be found in both the peripheral and central auditory pathways.

1.4.1. Peripheral auditory function

In terms of overall sex differences in the cochlear function, women have been shown to have better (more sensitive hearing) PTA thresholds than men across all frequencies (0.25–8 kHz) (Grinn et al., 2017). OAEs also show significant sex differences both related to their presence and strength (dB SPL) (McFadden, 1998). Specifically, women's cochleas are more likely to produce SOAEs than men's. The prevalence varies between studies, 70 vs. 60% (Penner and Zhang, 1997) or 85 vs. 45% (Talmadge et al., 1993), but the overall picture is similar. Snihur and Hampson (2011) found no sex effect in prevalence of SOAEs, but in the strength of SOAEs, with women having significantly stronger SOAEs than men. Burns et al. (1992) found significant sex differences in SOAE prevalence, not only in adults but also in neonates, with females having a higher number of SOAEs present than males. A potential explanation for these findings might be that in female neonates, umbilical cord blood at birth has higher levels of estradiol (Kuijper et al., 2013) than in male neonates. However, by the age of 24 months, these sex differences in SOAEs disappear, possibly because of decreased sex hormones levels in blood and the changes in the external and middle ears (Folsom et al., 1994). In terms of TEOAEs, sex differences have been shown for women (Burns et al., 1992; Shuster et al., 2019) and neonates (Burns et al., 1992; Newmark et al., 1997) with females having stronger TEOAEs than males, but not for older infants (Folsom et al., 1994; Newmark et al., 1997). Newmark et al. (1997) also found that there were fewer asymmetries recorded between both ears in women compared to men.

In contrast to SOAEs and TEOAEs, DPOAEs show no effect of sex hormones. The sex differences found in DPOAEs' phase delay (longer for men than women) can be fully explained by sex differences in the anatomical length of the cochlea rather than the differences in sex hormones (Bowman et al., 2000).

Sex differences in cochlear function may also contribute to differences in susceptibility to haring loss and in particular noise induces hearing loss. Estradiol can have protective role in the inner ear against noise exposure. Sex differences have been found in prevalence of noise-induced hearing loss (NIHL) between women and men (Pearson et al., 1995; Delhez et al., 2020; Wang et al., 2021). For instance, Wang et al. (2021) conducted a cross sectional study to investigate sex differences in NIHL among 2,280 industrial noise-exposed shipyard workers (1,140 women) and found that women were less likely to develop high-frequency hearing loss than men. It is important to note though that studying the effect of sex hormones on NIHL is particularly challenging in humans, as men are more likely to be exposed to excessive occupational noise than women. Therefore, matching the amount of noise exposure in the participants in order to give a clear view on the protective role of female hormones can be difficult.

Given the absence of progesterone receptors in the cochlea it is unlikely that progesterone has a direct effect on peripheral hearing in humans (Bonnard et al., 2013). However, progesterone receptors have been suggested to play an important role in the central auditory system by modulating the processing of auditory clues (Mann et al., 2012; Upadhayay et al., 2014).

Besides overall static differences in peripheral hearing function due to sex hormones, and particularly estradiol, dynamic sex differences have also been found during the ovarian cycle. Higher levels of estradiol (during the late follicular phase) have been suggested to be associated with a positive effect on hearing sensitivity as evaluated by audiometric threshold (Al-Mana et al., 2010). In particular, PTA thresholds have been reported to improve during higher levels of estradiol (Souza et al., 2017; Emami et al., 2018; Karaer and Gorkem, 2020). In addition, high levels of female hormones during the menstrual cycle have been found to increase the right ear advantage in women (Cowell et al., 2011; Carneiro et al., 2019) with significant differences being reported for cycle phases with high estradiol levels, i.e., follicular phase (Cowell et al., 2011).

As in human studies, animal studies have suggested a positive relationship between levels of estradiol and hearing sensitivity. During high levels of estradiol, better hearing responses were reported in female mice (e.g., Laugel et al., 1987; Canlon and Frisina, 2009; Frisina, 2012), fish (Sisneros et al., 2003, 2004), and frogs (Arch and Narins, 2009).

In addition, estradiol replacement therapy in ovariectomized rats results in a significant improvement in blood circulation in the cochlea (Laugel et al., 1987; Stenberg et al., 2003). This occurs possibly because estradiol inhibits ion transport from stria vascularis by enabling the ion channels in the stria vascularis to inactively secrete K^+ into the scala media, which in turn enhances the function of OHCs and IHCs (Lee and Marcus, 2001).

The pattern of systematic sex differences in SOAEs and TEOAEs but not DPOAEs has also been found in Rhesus monkeys (McFadden et al., 2005, 2006). McFadden et al. (2006) recorded OAEs in Rhesus monkeys prior to, during and post-breeding season. Female Rhesus monkeys showed stronger and more numerous SOAEs and TEOAEs than male Rhesus monkeys, with female TEOAEs being particularly high during the breeding season (higher estradiol and progesterone levels). No significant sex differences were found in their DPOAEs. There were also no differences in the DPOAEs during the breeding season when the differences in TEOAEs were highest (McFadden et al., 2006). As was already indicated by human studies DPOAEs do not appear to be sensitive to detecting sex differences or changes in cochlear function due to differences in hormone levels.

In studying the effect of progesterone on hearing, Price et al. (2009) found a significant reduction in hearing sensitivity in mid and high frequencies (in ABR and DPOAEs results) in ovariectomized mice that were treated with estradiol-progesterone hormone replacement therapy (HRT). The group of mice that were treated with estradiol monotherapy showed better results compared to the group that was treated with progesterone-containing HRT (Price et al., 2009).

Milon et al. (2018) studied the protective role of estradiol and sex differences in susceptibility to noise exposure in mice. They explosed male and female mice to 2 h of an octave-band of noise centered at 11.3 kHz (8–16 kHz), presented at 101 dB SPL, and found that female mice had significantly smaller permanent threshold shift at 16, 24, and 32 kHz than male mice. This result is in agreement with Meltser et al. (2008) who found that young female mice were more protected from acoustic trauma (12–25 dB threshold shift) than young males (15–26 dB threshold shift) and older female mice (32–49 dB threshold shift) across tested frequencies from 8 to 20 kHz.

1.4.2. Central auditory function

One measure of central auditory function, ABR, shows sex differences in its latencies and amplitudes of response. Specifically, pre-menopausal women have been shown to have larger amplitudes and shorter latencies ABRs (better ABRs) than age-matched men (Zakaria et al., 2019). McFadden (1998) and Meltser et al. (2008) showed that when levels of estradiol concentration in the inner ear were high, wave I latency of ABR decreased (indicating faster conduction) and the ABR amplitude increased, presumably because estradiol improves the neurotransmission of the acoustic signals. Systematic sex differences have been shown also for wave V of ABR, with shorter latencies and larger amplitudes in women than man. While the majority of these differences are thought to be due to differences in head size rather than hormones (women tend to have smaller heads compared to men resulting in a faster propagation of wave V and thus shorter latencies (Don et al., 1993). However, this anatomical difference cannot explain all differences between women and men. Don et al. (1993) showed that the sex differences remained in ABRs even when the size of participants' heads was considered, suggesting that there might be some role for hormones after all. This interpretation is supported by findings from Zakaria et al. (2019), who recorded ABRs at supra-threshold and threshold levels in young adults from both sexes, while considering comparative head size. They found consistent sex differences in the ABRs with women having better responses (shorter latencies and higher amplitudes of ABRs) at the supra-threshold levels than men (Zakaria et al., 2019).

The relationship between sex differences and speech perception is rarely mentioned in the literature and information on the differences between sexes in speech perception is limited. Wadnerkar et al. (2008), using consonant-vowel (CV) syllable perception to study sex differences in dichotic listening, reported sex differences in dichotic listening asymmetry at lower estradiol levels, but not at higher estradiol levels. Using dichotic digits, staggered spondaic word, and dichotic consonant-vowel tests to study dichotic listening during the menstrual cycle, Carneiro et al. (2019) found sex differences during periods of high levels of estradiol in women when compared to a control group of men. Specifically they found that the right ear in women (compared with the left ear and test session in men) significantly differs during periods of high levels of estradiol in staggered spondaic word and dichotic consonant-vowel tests, but not in dichotic digits (Carneiro et al., 2019).

Sex differences in overall hearing function are further complicated by short-term fluctuations in sex hormones that occur during the ovarian cycle. Dehan and Jerger (1990) reported changes in ABRs that occurred in synchrony with monthly changes in female sex hormones, indicating a possible influence of cyclical sex hormone fluctuations on latencies of ABRs. The nature of this influence is still unclear, with some recent studies suggesting that the effect of estradiol on ABRs may be negative, such that high levels of estradiol prolong latencies of ABRs (Disney and Calford, 2001; Al-Mana et al., 2010).

Fewer studies on the cyclical effect of hormones on hearing in animals are available. Sisneros et al. (2003) investigated cyclical changes in hearing and found that during breeding seasons (higher estradiol levels) the auditory nerve of female Midshipman fish (who have a vocal form of breeding) showed an increase in response to male mating fish. Moreover, when female midshipman fish were treated with estradiol during non-breeding seasons (lower estradiol levels), it resulted in an increase in the sensitivity of their auditory nerve (Sisneros et al., 2004).

The only one study suggesting that progesterone receptors may play an important role in the central auditory processing, and specifically in modulating the processing of auditory clues, comes from túngara frogs (O'Connell et al., 2011). O'Connell et al. (2011) measured auditory activities in anterior, lateral, and ventral thalamic nuclei, as these regions contain progesterone receptors. It was found that progesterone may act as a processing modulator of the auditory inputs. In addition, progesterone receptors were found in both the striatum and medial pallium in this species, which provides another path of progesterone modulation of the auditory input.

1.5. The effect of reduced levels of estradiol on auditory function

In both animal and human studies, estradiol has been reported to have multiple protective properties in the inner ear (Mitre et al., 2006; Tremere et al., 2009) and to contribute to protecting the ear from noise exposure, to delay the onset of age-related hearing loss, and to aid spontaneous recovery from sensory-neural hearing loss (e.g., Köşüş et al., 2012; Zhang et al., 2018; Delhez et al., 2020). Therefore, reduced levels of estradiol may cause hearing loss, in particular in menopause, and in Turner Syndrome.

1.5.1. Menopause

Sex differences related to menopause have been reported in terms of onset and severity of age-related hearing loss (ARHL). While men develop ARHL before age-matched women (Davis, 1995), women experience a faster decline in hearing than men after menopause (Hederstierna et al., 2010; Villavisanis et al., 2018). Indeed, the earlier reported advantage in hearing sensitivity for pre-menopausal women compared to men reverses with age (McFadden, 1993) with older women having worse thresholds (i.e., worse hearing sensitivity) than age-matched men (Corso, 1968; Mościcki et al., 1985; Jerger et al., 1993; Kim et al., 2010).

While it has been suggested that hormonal changes in menopause may cause ARHL in post-menopausal women (Wharton and Church, 1991), the actual mechanism of the effect of the reduced levels of hormones on hearing sensitivity for this age group is unclear. Evidence that lower levels of estradiol may play a critical role comes from Karaer and Gorkem (2020) who reported no differences in hearing between pre-menstrual women with premature ovarian failure and post-menopausal women. Similarly, Kim et al. (2002), who studied the association of serum estradiol levels and hearing sensitivity in post-menopausal women, found that lower levels of estradiol increased the risk of hearing loss. In addition, Arora et al. (2021) compared the ABRs of post-menopausal women and pre-menopausal women. They reported that post-menopausal women had significantly reduced amplitudes and prolonged latencies of ABRs. On the other hand, non-significant differences in ABRs using sensation level as stimulus between older men compared to young men found by Anias et al. (2004). Rosenhamer et al. (1980) found non-significant differences in ABRs between post-menopausal women compared to age-matched men and young men. This may concluded that sex hormones may influence ABRs.

Recent attention has focused on the effect of hormone replacement therapy for improving hearing ability in post-menopausal women. According to studies that examined the connection between female sex hormones and hearing, hormonal treatments tend to delay hearing loss in post-menopausal women (Kilicdag et al., 2004; Köşüş et al., 2012; Lee et al., 2017). Furthermore, post-menopausal women who undertook hormonal therapy had better hearing sensitivity than women who did not take HRT. Kilicdag et al. (2004) studied two groups of postmenopausal women, where only one group was given estrogen treatment. They reported that hearing sensitivity at 250–2,000 Hz was better in the group who had the estrogen treatment compared to the control group. In addition, Caruso et al. (2003), when investigating auditory function of women with induced early menopause due to medical intervention, found a decline in auditory function. However, with low doses of estrogen treatment, hearing function improved as demonstrated by shortened latencies of ABRs (Caruso et al., 2003). Even though estrogen hormonal treatment could be a novel approach to restoring and delaying hearing loss, there is a controversy regarding its potential for increasing the risk of developing breast cancer. A randomized placebo-controlled study, however, showed that only a estrogen and progestin combined HRT increased the risk of breast cancer, while an estrogen-only HRT significantly decreased the risks (Chlebowski et al., 2020). In addition, a recent review found that estrogen HRT not only has the potential to prevent breast cancer, but may also be able to help prevent other disorders (e.g., osteoporosis and cardiovascular disease) (Manyonda et al., 2022).

1.5.2. Turner syndrome

Turner syndrome represents another example of the consequences of lack of estradiol for hearing. Turner syndrome is a genetic condition in women caused by either complete or partial deletion of the X chromosome that leads to ovarian dysgenesis and little or no estradiol production. Turner syndrome has been associated with low level of estradiol, which in turn has been suggested to play a critical role in development of hearing impairment (Morimoto et al., 2006; Hederstierna et al., 2009).

As secretion of female sex hormones only starts to increase in girls with puberty, and lack of secretion and resulting hearing loss is normally only detected after girls start puberty. Stenberg et al. (1998) reported that young girls with Turner syndrome had within-normal hearing levels before puberty (i.e., as the peak of the sensorineural dip did not exceed 20 dB HL), however, their hearing sensitivity to high frequencies decreased after puberty (i.e., the dip between 25 and 35 dB HL).

1.5.3. Pregnancy

As mentioned earlier, changes in the circulating levels of female sex hormones may affect the functioning of the auditory system. This is also seen during pregnancy (Sennaroglu and Belgin, 2001). Progesterone is considered the main sex hormone during pregnancy, as it is essential in fetus implementation and pregnancy maintenance (Di Renzo et al., 2005). The production levels of progesterone increase significantly during pregnancy (from 0.1 to 40 mg/24 h in non-pregnant women to 250–600 mg/24 h in near-term pregnant women; Sennaroglu and Belgin, 2001).

There is some evidence that low-frequency hearing is slightly elevated in pregnant women, particularly in the third trimester (Sennaroglu and Belgin, 2001). This is most likely due to fluid retention in the inner ear. However, this elevation tends to remain within normal clinical levels. In addition to changes in hearing sensitivity, reduction in DPOAE have been found with DPOAE being absent in 26% of the pregnant as opposed to 4% of the non-pregnant women (Ashok Murthy and Krishna, 2013).

A possible explanation for these changes in hearing function is the substantial increase in progesterone levels in pregnancy, which can lead to edema. Edema can have an effect similar to endolymphatic hydrops in the cochlear aqueduct and essentially lead to a temporary conductive loss. During the post-partum period, when progesterone is reduced again, hearing levels have been found to spontaneously recover (Sennaroglu and Belgin, 2001; Kenny et al., 2011).

1.6. Contribution to the field

Sex affects hearing function, yet its effect is regularly ignored. This changed only in 2016 when in the UK sex was added as a biological variable in preclinical research by the National Institutes of Health (Clayton and Collins, 2014); in earlier studies, sex was commonly not reported and analyzed separately, opening the door to the possibility that existing sex differences in the data were not discovered and may have inadvertently affected the results. Based on the results reviewed so far, this bias is most likely to have affected studies of ARHL and NIHL.

Besides a general lack of focus on sex differences in hearing research, there is also the problem of comparability of methodologies for measuring hearing function and hormone levels. One case in point is measures used assess the point of the menstrual cycle. While the most accurate method would be to measure hormone levels in the blood of the participant at the point of auditory testing, most previous work has used self-reported measures. However, we know that self-report measures are less accurate that biological assessment at estimating levels of estradiol in the bloodstream (Farrar et al., 2015), yet studies using either method are treated as comparable. This can make is difficult to develop a clear understanding of how sex hormones affect the various stages of the auditory pathway in an overall and cyclical fashion.

A lack of accuracy in measures and consistency and reliability between measures makes it difficult to combine data from existing studies into in a meta-analysis to obtain a clearer picture of the effects of sex hormones on hearing. As a result, there has been no systematic review of the differences in the auditory function between women and men, and the effect of female hormones fluctuations on auditory function across a specific period (i.e., during menstrual cycle and after menopause) to date. In addition, the possible effect of the female hormones on auditory dysfunction, such as perception of tinnitus and vestibular dysfunction, is unclear. Therefore, this review aims to systematically assessed the literature to give an overview of the methodologies used in research and to identify the gap in knowledge and to make some recommendations for future work to have a better understanding of the association between the levels of these hormones and hearing. This will provide information about how to manage hearing loss, tinnitus, and vertigo in women.

This systematic review aims to answer the following questions:

> Review question 1. Does auditory function differ between women and men across the entire lifespan or during part of it?
> Review question 2. Does auditory function in women fluctuate over the course of the menstrual cycle?
> Review question 3. Does this fluctuation co-vary with changes in female hormone levels?

2. Review methods

The protocol of this review was registered with the International Prospective Register of Systematic Reviews (PROSPERO; Reference ID: CRD42020201480) in October 2020 (NIHR, 2020).

https://www.crd.york.ac.uk/prospero/display_record.php?ID~=~CRD42020201480. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines were used to formulate the eligibility criteria (Moher et al., 2015; Page et al., 2021).

2.1. Eligibility criteria

2.1.1. Participants

- Studies of pre-menopausal women/and age-matched men with normal hearing.
- Studies of pre-menopausal women with a regular menstrual cycle, no use of hormonal contraceptives, no pregnancy, and no lactation.
- Studies of post-menopausal women/and age-matched men with normal hearing/hearing loss.

2.1.2. Intervention/Exposure

Estradiol and progesterone.

2.1.3. Comparators

If reported, age-matched men.

2.1.4. Outcomes

Measures of peripheral and central auditory function. Peripheral auditory function: pure- tune audiometry (PTA, conventional and extended high frequencies), tympanometry, medial olivocochlear reflex (MOC) and otoacoustic emissions (OAEs); central auditory function: auditory brainstem responses (ABR), auditory steady state responses (ASSR), speech audiometry, auditory evoked/event-related potentials (AEP and ERP) and any further recommended procedure (e.g., dichotic listening).

2.1.5. Study designs

Randomized controlled trials, non-randomized controlled trials, case- control, cross-sectional and/or prospective cohort/longitudinal studies.

2.2. Information sources

The following electronic databases were searched (EMBASE, PubMed, MEDLINE (Ovid), PsycINFO, ComDisDome, CINAHL, Web of Science, and CENTRAL via Cochrane Library). Additional to the electronic databases, reference lists of relevant studies and reviews were scanned, and relevant internet resources (e.g., Mendeley) were searched for relevant publications published between 1999 and 2022. The search strategy is in the **Supplementary material**.

2.2.1. Selection process

The title and abstract were screened for all retrieved articles. The eligibility of the retrieved articles was assessed according to the inclusion/exclusion criteria by NA. In case of uncertainty, this was solved by discussion with AH and KK.

2.2.1.1. Inclusion criteria

- Published studies in English, or if English translation was available.
- Studies done on human participants, adults ($\geq$17 years).
- Only human participants were included:
  - Pre-menopausal women.
  - Post-menopausal women.
  - Adult men.

2.2.1.2. Exclusion criteria

- Gray literature, systematic review, conference abstracts, book chapters, dissertations, theses, and clinical guidelines.
- Pre-clinical studies/Animal studies.
- Studies that included female participants who were breastfeeding, pregnant or the use contraceptive pills or if not mentioned.
- Studies including participants with additional health conditions or risk factors for ototoxicity, noise exposure and middle ear pathologies.

2.3. Data management

The identified papers were extracted to EndNote X9 (Clarivate Analytics, 2018) for the initial screen. Duplicates were removed prior to the screen using the same software. The reviewer NA transferred the following information into an Excel spreadsheet: the titles, authors' names, year of publication, settings, participants characteristics, publication journals, study design, abstracts, number of sessions, outcome measures (including hormones levels measures), and findings. The excluded papers were documented in the spreadsheet with the reason for exclusion.

2.4. Risk of bias assessment

The risk of bias of each individual study was assessed using Newcastle-Ottawa Scale (NOS). This scale judges the quality of papers in three broad perspectives: the selection, comparability, and outcome. In addition, NOS assess the following: control cohort, the number of session (the length/follow up), and outcomes measures (objective or self-reported). The quality of the studies could be judged as either good (low risk), fair (high risk), or poor (very high risk) by awarding stars in each domain accordingly with NOS guidelines (Wells et al., 2000).

2.5. Data analysis

The information collected from the systematic review was analyzed qualitatively and represented in tables and paragraphs form. Such material would include participant characteristics, test criteria, outcome measures, and findings. This review did not use meta-analysis due to the amount of the missing data (i.e., SD, and number of participants in each group were not reported). Attempts were made to contact the author(s), but this information was not provided. Therefore, each paper was assessed to reach a general conclusion.

2.6. Search results

The initial search of the databases recorded 6,958 articles. 173 articles were duplicates and removed by automation tools. After removal of duplicates, titles were screened to identify relevant studies. The screening identified 6,732 potential articles. An additional 17 articles identified through Mendeley, and hand search (i.e., checking references and citation). After screening the titles, 165 articles remained for abstract screening. After abstract screening, 119 articles were excluded. The full text assessment of the remained 48 articles resulted in identifying 35 articles that meet the inclusion criteria. A summary of the selection process is presented in the PRISMA flow chart diagram (**Figure 2**).

2.7. Study characteristics

The included studies were divided into three groups based on participant characteristics and study design: studies on the

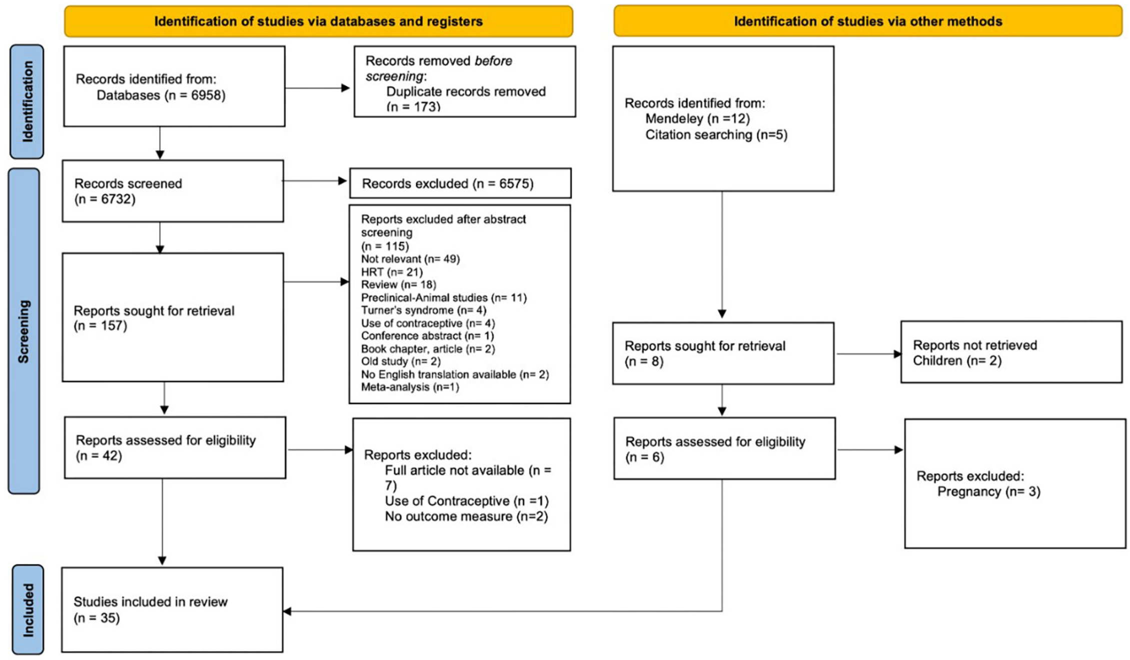


FIGURE 2
Preferred reporting items for systematic reviews and meta-analyses 2020 selection process flow.

sex differences between pre-menopausal women and age-matched men in auditory function, studies on the female hormones' fluctuation in pre-menopausal women, and auditory changes in post-menopausal women.

Changes in levels of female sex hormones were measured using biological samples (i.e., blood or saliva, used in 11 studies) or self-reported measures (i.e., day counting, used in 7 studies) to predict the phase of the menstrual cycle and then infer the level of female sex hormones. This was done by counting the day of the cycle according to the participants' average menstrual cycle length (i.e., this is calculated from the first day of last menstrual period). Six studies were unclear on the methods used to measure the female hormones.

2.7.1. Sex differences between pre-menopausal women and age-matched men

Eleven studies investigated sex differences in the peripheral and central auditory system (Bowman et al., 2000; Ismail and Thornton, 2003; Dreisbach et al., 2007; Sharashenidze et al., 2008; Kim et al., 2010; Snihur and Hampson, 2011; Jalaei et al., 2017; Boothalingam et al., 2018; Melynyte et al., 2018; Stuart and Kerls, 2018; Zakaria et al., 2019). A summary of the characteristics of the studies are presented in **Table 1**.

The sample size ranged from 29 to 1,116 participants. Eight studies had similar participant characteristics; young adults with normal hearing (Bowman et al., 2000; Ismail and Thornton, 2003; Dreisbach et al., 2007; Snihur and Hampson, 2011; Jalaei et al., 2017; Melynyte et al., 2018; Stuart and Kerls, 2018; Zakaria et al., 2019). Three studies included young and older participants. In the Boothalingam et al. (2018) all participants had normal hearing; however, in Sharashenidze et al. (2008) and Kim et al. (2010), older participants reported to have ARHL. In addition, for older participants no history of excessive noise exposure was reported in these studies. Six studies reported OAE outcomes, three PTA, 2 ABR, and one ASSR outcomes. Most outcomes were reported for young adults (eight studies), but three studies reported outcomes across the whole adult age range. In all cases the studies only comprised one testing session.

2.7.2. Female hormone fluctuations in pre-menopausal women

Nineteen articles studied the effect of female sex hormone fluctuations on auditory function throughout the auditory pathway in women across the menstrual cycle (Serra et al., 2003; Yadav et al., 2003; Walpurger et al., 2004; Wadnerkar et al., 2008; Al-Mana et al., 2010; Cowell et al., 2011; Hjelmervik et al., 2012; Mann et al., 2012; Griskova-Bulanova et al., 2014; Upadhayay et al., 2014; Hodgetts et al., 2015; Adriztina et al., 2016; Batta et al., 2017; Hu and Lau, 2017; Liu et al., 2017; Souza et al., 2017; Emami et al., 2018; Carneiro et al., 2019; Karaer and Gorkem, 2020). A summary is presented in **Table 2**.

The sample size of these studies ranged from 16 to 94 participants. The participant characteristics are similar in seventeen studies: young adult, normal hearing levels, regular menstrual cycle, no pregnancy, no lactation for 6 months prior to testing. One study included women with premature ovarian failure and normal hearing levels (Karaer and Gorkem, 2020). One study did not mention the regularity of the menstrual cycle of female participants (Liu et al., 2017).

Only seven studies had a control group. In addition, the make-up of control groups differed between studies. In most studies (five), the control were age-matched men (Wadnerkar et al., 2008; Hjelmervik et al., 2012; Liu et al., 2017; Souza et al., 2017; Carneiro et al., 2019). Three studies either exclusively or additionally used control groups comprised of women with premature ovarian failure cycle and older women with ARHL (Karaer and Gorkem, 2020) or women who use a method of hormonal contraception as a control (Yadav et al., 2003; Souza et al., 2017).

The number of sessions varied in these studies. Testing sessions varied between four (Yadav et al., 2003; Al-Mana et al., 2010; Mann et al., 2012; Hu and Lau, 2017; Souza et al., 2017), three (Serra et al., 2003; Walpurger et al., 2004; Hjelmervik et al., 2012;

TABLE 1 Summary of the characteristics and results of studies on "sex differences in auditory function."

References	Study design	Sample size	Age mean	Hearing level	Focus	Outcome measures	Findings
Bowman et al., 2000	Comparative study	Women (30) Men (30)	19-35 years women (25.0 years) men (25.6 years)	Normal hearing	Sex differences in DPOAE's	DPOAE recordings	There are sex differences in DPOAEs recordings, but these differences are related to the anatomical differences in cochlear length between sexes, not differences in hearing sensitivity. At low frequencies, men had longer DPOAE measures than women.
Ismail and Thornton, 2003	Comparative Study	Women (40) ears Men (41) ears	18-40 years	Normal hearing	Sex differences in MLS OAE	MLS OAEs recordings	There are sex differences in MLS OAEs. Women had a greater MLS OAEs amplitude than men. The relevance of this difference, however, diminishes as the click stimulation rate increases. Women's right ears reported to have greater MLS OAEs amplitude of than women's left ears.
Dreisbach et al., 2007	Comparative Study	Women (30) Men (30)	18-39 years	Normal hearing	Race and sex differences in PTA and DPOAEs	PTA DPOAEs	Women had better hearing sensitivity at 14 K and 16 KHz. No racial or sex differences were found for the DPOAE measure.
Sharashenidze et al., 2008	Comparative study	Women (128) Men (96)	30-79 years	Hearing levels varied among the age groups and sex	Sex differences in age-related hearing loss/presbycusis	PTA	Women had better hearing levels than men in age groups of 30-39, 40-49 and 50-59 years. In the age group of 60-69 and 70-79 years, women tend to have a steeper decrease in hearing, and the sex differences in hearing sensitivity are smoothed significantly.
Kim et al., 2010	Comparative study	Women (902) Men (214)	15-83 years women (46 years) men (47.6 years)	Young group: normal hearing Old group: ARHL	Sex differences in ARHL	PTA	There are significant sex differences in PTA thresholds. Women have better hearing at higher frequencies than men. At 4 kHz and 8 kHz, men reported to have greater age-related changes in hearing than women.
Snihur and Hampson, 2011	Comparative study	Women (48) Men (45)	17-25 years women (19.9 years) men (20.8 years)	Normal hearing	Sex differences in SOAE and CEOAE	SOAE CEOAE recordings	There are sex differences in SOAEs and CEOAE. Women producing more numerous and stronger SOAEs, and CEOAEs with greater response amplitude compared to men.
Jalaei et al., 2017	Comparative study	Women (15) Men (14)	19-30 years Women (23.5 years) Men (22.7 years)	Normal hearing	Sex differences in speech-ABR	Speech-ABR	Significant sex differences in the amplitude of speech-ABR peaks V and A. Higher amplitudes and less steep V/A slopes were observed in women than in men, and these differences persisted when considering differences in head size. Women were found to have shorter latencies of peak V and A. However, the differences in latencies were insignificant when considering the differences in head size.
Boothalingam et al., 2018	Comparative study	Women (522) Men (365)	10-68 years	Normal hearing	Sex, race, ear differences in DPOAE's	DPOAE recordings	There are no significant sex differences in DPOAE recordings found in the study.
Melynyte et al., 2018	Comparative study	Women (22, 11 left-handed) Men (22, 11 left-handed)	left-handed (23 years) right-handed (22 years)	Normal hearing	Sex differences and handedness in 40 Hz ASSR	40 Hz ASSR	There are sex differences observed in the left-handed participants, as women significantly had lower phase-locking and event-related spectral perturbation values of 40 Hz ASSRs compared to the left- handed men. However, no significant sex differences between right-handed women and men.

(Continued)

TABLE 1 (Continued)

References	Study design	Sample size	Age mean	Hearing level	Focus	Outcome measures	Findings
Stuart and Kerls, 2018	Comparative study	Women (50) Men (50)	Women (22.1 years) Men (23.4 years)	Normal hearing	Sex differences in contralateral inhibition of transient evoked otoacoustic emissions (TEOAEs)	Contralateral TEOAEs	There are significant sex differences in TEOAEs recording. The levels of TEOAEs were larger in women and in the right ear than in men and the left ear. There is no significant effect of ear or sex on absolute TEOAEs inhibition. Significant negative correlations and linear predictive relations were found between TEOAE levels and normalized TEOAE inhibitions in both ears. There is no evidence of the same with absolute inhibition of TEOAEs. The effect of ear and sex on normalized inhibition are small and may have no clinical or practical significance.
Zakaria et al., 2019	Comparative study	Women (17) Men (13)	Women (22.6 years) Men (21.9 years)	Normal hearing	Sex differences in ABR at Suprathreshold	ABR	A significant sex differences in ABR results among young adults were found at suprathreshold levels. These differences are not related to the head size. Normative data for sex differences in ABR are valuable for clinical applications, particularly at high stimulation levels.

Adriztina et al., 2016; Batta et al., 2017), two (Wadnerkar et al., 2008; Upadhayay et al., 2014; Emami et al., 2018; Carneiro et al., 2019), and one (Cowell et al., 2011; Griskova-Bulanova et al., 2014; Hodgetts et al., 2015; Liu et al., 2017; Karaer and Gorkem, 2020). In studies with more than one session, there was at least an element of repeated testing to compare outcome measures within participants across different phases of the menstrual cycle. In studies with only one session, all comparisons between outcome measures across different phases of the menstrual cycle were between-participant.

2.7.3. Auditory changes in post-menopausal women

Five studies investigated auditory changes in post-menopausal women, with the sample size ranging from 22 to 190 participants (Tandon et al., 2001; Hederstierna et al., 2010; Svedbrant et al., 2015; Trott et al., 2019; Arora et al., 2021). One study tested participants three times (i.e., at 2, 7, and 10 years after the start of menopause) (Svedbrant et al., 2015). One study tested participants twice with mean interval of 7.5 years between the two sessions (Hederstierna et al., 2010). Three studies tested participants once (Tandon et al., 2001; Trott et al., 2019; Arora et al., 2021). A summary is presented in **Table 3**. None of these studies included age-matched men as a control. In Hederstierna et al. (2010) and Svedbrant et al. (2015) participants had a baseline normal hearing level. In addition, the participants in Tandon et al. (2001), Trott et al. (2019), and Arora et al. (2021) studies reported to have normal hearing level.

2.8. Outcomes measures

2.8.1. Audiometric measures

Sex differences between pre-menopausal women and age-matched men were assessed across the peripheral auditory pathway using PTA (Dreisbach et al., 2007; Sharashenidze et al., 2008; Kim et al., 2010), SOAEs and COAEs (Snihur and Hampson, 2011), maximum length sequence OAEs (MLS OAEs) (Ismail and Thornton, 2003), contralateral TEOAEs (Stuart and Kerls, 2018), and DPOAEs (Bowman et al., 2000; Dreisbach et al., 2007; Boothalingam et al., 2018). In addition central auditory measures assessed ABR (Zakaria et al., 2019), speech-ABR (Jalaei et al., 2017), and 40 Hz ASSR (Melynyte et al., 2018).

Fluctuations in female hormones in pre-menopausal women were assessed using the following peripheral auditory measures: middle ear function (Adriztina et al., 2016; Emami et al., 2018), TOAEs (Al-Mana et al., 2010; Karaer and Gorkem, 2020), DPOAEs (Adriztina et al., 2016; Emami et al., 2018; Karaer and Gorkem, 2020), PTA (Adriztina et al., 2016; Souza et al., 2017; Emami et al., 2018; Karaer and Gorkem, 2020) and medial olivocochlear suppression (Al-Mana et al., 2010). Central auditory function was assessed using dichotic speech audiometry (Wadnerkar et al., 2008; Cowell et al., 2011; Hjelmervik et al., 2012; Hodgetts et al., 2015; Hu and Lau, 2017; Carneiro et al., 2019), ABR (Serra et al., 2003; Al-Mana et al., 2010; Mann et al., 2012; Upadhayay et al., 2014; Batta et al., 2017; Hu and Lau, 2017; Emami et al., 2018), long latency auditory evoked potentials (LLEAPs) (Yadav et al., 2003), 40 Hz auditory steady-state response (ASSR) (Griskova-Bulanova et al., 2014), event-related potentials (ERPs) (Walpurger et al., 2004), and speech-ABR (Liu et al., 2017).

Auditory changes in post-menopausal women were assessed using PTA (Hederstierna et al., 2010; Svedbrant et al., 2015), dichotic digit test, Speech in noise tests, middle latency response (MLR) (Trott et al., 2019) and ABR (Tandon et al., 2001; Trott et al., 2019).

A summary of the hormonal tests for each study that investigated female hormones fluctuation in pre-menopausal women is presented in **Table 2**. Only one study measured female hormone levels in post-menopausal women, and they used blood samples (Svedbrant et al., 2015). The method of measuring the level of the hormones was not mentioned in Tandon et al. (2001) and Hederstierna et al. (2010).

TABLE 2 Summary of the characteristics and results of studies on "the fluctuation of auditory function during the menstrual cycle."

References	Study design	Sample size	Age mean (SD)	Hearing level	Control group	Number of sessions	Experimental group description	Outcome measures		Findings
								auditory	hormones	
Serra et al., 2003	Observational study	Women (94)	27.9 (6.1)	Normal hearing	No control	Three sessions Early follicular phase (day 5-8) Late follicular phase (day 13-16) Early luteal phase (day 18-23)	Regular cycle (28.3, SD 3.3)	ABR	Enzyme-linked immunosorbent assay	Shorter wave latencies and interpeak intervals during the late follicular phase than during the early luteal phase.
Yadav et al., 2003	Observational study	Women (40) [20 women use contraceptive pills (CP)]	19-26	Normal hearing	age-matched women taking hormonal contraceptive pills	Four sessions in a single cycle 1. Early follicular phase (day 1-3) 2. Late follicular phase (day 11-14) 3. Early luteal phase (day 17-22) Late luteal phase (day 25-27)	Regular menstrual cycles Anovulatory cycle/use of contraceptive pills	LLAEPs	Day counting	P2 and N2 latencies varied significantly throughout the phases of the cycle in normal cycling women. The latencies increased from early to late follicular phase and decreased during early luteal phase and increased again in late luteal phase. Similar but insignificant changes in P1 and N1 were observed. No changes or variation were noticed in CP group, LLAEPs remained consistent.
Walpurger et al., 2004	Observational study	Women (18)	18-35 years 26.5 (5.7)	Normal hearing	No control	Three sessions 1. Early follicular phase 2. Late follicular phase 3. late luteal phase	Regular cycle (24-35 days) No use of contraceptive pills	Event-related potentials (ERPs)	Saliva sample	There are changes in auditory ERPs across the menstrual cycle. The most prominent changes were observed during the late luteal phase, where the vertex potential was significantly reduced compared to menses and to the follicular phase. Which suggests that during high estradiol and progesterone levels in the luteal phase, the involuntary cortical arousal response to the external stimuli is reduced.
Wadnerkar et al., 2008	Observational study	Women (25) Men (20)	Women: 22.56 (2.04), Men: 22.15 (1.69)	Normal hearing	Age-matched men	Women tested in two sessions during one cycle: Early follicular phase (day 2-5) Between two phases, the early and late luteal phase (day 18-25) Men tested once	Regular menstrual cycle (29.24 days, SD2.45),	Dichotic CV stimuli	Day counting	No significant effect of the menstrual cycle on dichotic listening. Number of responses did not differ between the groups.
Al-Mana et al., 2010	Observational study - longitudinal	Women (16)	31.4 (8)	Normal hearing	No control	Four sessions Early follicular phase (5-8 days) Late follicular phase (10-14 days) Early luteal phase (20-23 days) Late luteal phase (25-28 days)	Regular cycle (28.5, SD 1.6)	SOAEs TEOAEs MOC suppression, ABR	Blood samples	During late follicular phase, SOAE amplitudes were significantly greater. The linear regression analysis of all TEOAEs in four sessions showed no correlation with E2. However, In the early and late follicular phase, positive correlation between TEOAEs and E2 was reported, and negative correlation between MOC and E2. The regression analysis of the correlation between TEOAEs and MOC and progesterone level showed no significant findings. ABRs showed a significant change during the ovarian cycle, with an increase in the wave V latency in the late follicular phase and a decrease in the early and late luteal phase.

(Continued)

TABLE 2 (Continued)

References	Study design	Sample size	Age mean (SD)	Hearing level	Control group	Number of sessions	Experimental group description	Outcome measures		Findings
								auditory	hormones	
Cowell et al., 2011	**Observational study-cross-sectional**	**Women (21)**	**25.24 (0.74)**	**Normal hearing**	**No control**	**One session** **8 women started in the early follicular phase, 2 in the late follicular phase, and 11 in the early luteal phase.**	**regular cycle (29.20, SD 0.96)**	**CV dichotic tests**	**Blood sample**	**Sex differences in dichotic listening found to be dependent to the hormonal status in women.** **Increases in the right ear advantage (REA) were found in women throughout periods of the menstrual cycle. REA was greater during higher levels of ovarian hormone.** **Left ear scores decreased during higher levels of luteinizing hormones (LH).**
Hjelmervik et al., 2012	Randomized Controlled Trial	Women (15) Men (15)	Women 23.5 years (5.1) Men: 23.1 years (2.4)	Normal hearing	Age-matched men	Three sessions for both groups For women: Early follicular phase (day 2-4) Late follicular phase (day 8-12) Early luteal phase (day 20-22)	Regular cycle (26-32 days)	Dichotic testing: CV	Saliva sample	Women perform better in the late follicular phase compared to the early follicular and early luteal phases.
Mann et al., 2012	Observational study-cross-sectional	Women (50)	19-36 years	Normal hearing	No control	Four sessions Early follicular phase (day 1-3) Late follicular phase (day 11-14) Early luteal phase (day 17-22) Late luteal phase (day 25-27)	Regular menstrual cycles (28-30 days) and they had not taken any hormonal pills during the past 6 months.	ABR	Day counting	During the late follicular phase, the waves latencies were increased, that showed a slower neural conduction. This can be attributed to the high levels of estradiol during the late follicular phase of the menstrual cycle. The waves latencies decreased in the early luteal phase and hence, this enhanced the conduction across the neural pathways.
Griskova-Bulanova et al., 2014	Observational study-cross-sectional	Women (28)	20.68 years (0.63)	Normal hearing	No control	During one of the cycle phases Early follicular phase Late follicular phase Early luteal phase	Regular cycle (28.59, SD 2.13)	40 Hz ASSR	Saliva sample	Significant effect of menstrual cycle phase was seen for the total intensity of 40 Hz ASSRs. ASSR amplitudes were highest during the late follicular phase, intermediate during the early follicular phase and lowest during the early luteal phase. No relationship of any measures to progesterone concentrations was observed.
Upadhayay et al., 2014	Observational study-cross-sectional	Women (40)	16-26 years mean: 19 years (2.35)	Normal hearing	No control	Two sessions (one session during the follicular phase and another session during the luteal phase) The testing sessions were reported to be between 2-4 days before ovulation and 9-11 days after ovulation, according to their menstrual cycle. The exact testing days were not reported.	Regular cycle, No use of contraceptive pills, no pregnancy, no lactation	ABR	Day counting	There was a significant variation in ABR waves in the menstrual cycle. Better ABR recordings were observed during luteal phase compared to follicular phase of menstrual cycle. Progesterone is the likely hormone responsible for the increase in the conduction of auditory pathways in women of reproductive age group.
Hodgetts et al., 2015	Randomized Controlled Trial	Women (73)	23 years (4.86)	Normal hearing	No control	One session: The testing day was selected according to the women's self-reported cycle day (days 1-4, 7-12, 15-23, corresponding to the menstrual, follicular, or luteal phase, respectively)	Regular menstrual cycle (24-35)	Dichotic CV	Saliva sample	High levels of estradiol reported to reduce the stimulus-driven (bottom-up) aspect of lateralization rather than top-down cognitive control.

(Continued)

TABLE 2 (Continued)

References	Study design	Sample size	Age mean (SD)	Hearing level	Control group	Number of sessions	Experimental group description	Outcome measures		Findings
								auditory	hormones	
Adriztina et al., 2016	Observational study-cross-sectional	Women (49)	20-40 years	Normal hearing	No control	Three sessions during one cycle: Early follicular phase (day 3). Late follicular phase: tested with the ovulatory kit, indicating the estradiol at a high level. Early luteal phase: (day 21-22).	Regular menstrual cycle (24-35)	PTA, Tympanometry DPOAEs	Day counting	There was no significant correlation between menstrual and hearing thresholds. It was reported that during late follicular phase, there was a reduction in hearing sensitivity at 4 kHz in the right ear. However, DPOAEs amplitude significantly increased during late follicular phase, compared to early follicular and early luteal phase. This might suggest a positive effect of high levels of estradiol on the cochlear function.
Batta et al., 2017	Observational study-cross-sectional	Women (80)	18-24 years (18.8)	Normal hearing	No control	Three sessions during one cycle Early follicular phase (day 1-3) Late follicular phase (day 10-12) Early luteal phase (day 20-22)	Regular menstrual cycle and no use hormonal contraceptives	ABR	Not mentioned	There is a decrease in waves latencies during late follicular phase. It was reported that estradiol increases transmission in the auditory pathways, and it might be responsible for the shorter latency values of ABR. However, this variation is not statistically significant. There is no effect of female sex hormones on ABR waves amplitudes.
Hu and Lau, 2017	Observational study - longitudinal	Women (20)	21.5 (0.8)	Normal hearing	No control	Four sessions: Early follicular phase (day 3-4). Late follicular phase (day 9-10). Late follicular phase (ovulation) (day 14-15). Early luteal phase (day 21-22).	Regular cycle	ABR CV	Day counting	Peak V latency reported to be significantly lengthened during late follicular phase, but it is not true at peak I and peak III. The interpeak conduction times of inter-peaks I-V and III-V were prolonged at late follicular phase. It was concluded that the central conduction time depends on the phase of the menstrual cycle, which might affect dichotic listening performance.
Liu et al., 2017	Comparative study	Women (17) Men (18)	24-34 years Women (27.29 years) Men (28.17 years)	Normal hearing	Age-matched men	One session, the day of the cycle was not reported. the levels of estradiol and testosterone concentration were measured after the testing session.	Not mentioned	Speech-ABR	Blood samples	Significant effect of sex hormones on speech encoding in the brainstem. Estradiol is observed to affect the amplitude of neurons but has little effect on the conduction velocity of neurons (latency). Estradiol may improve brainstem auditory neuron excitability and phase-locking ability for speech coding.
Souza et al., 2017	Comparative study	**Women (20) Men (10)**	**18-39 years**	**Normal hearing**	**Age-matched men**	**Four sessions over one cycle Early follicular phase (day 1-7) Late follicular phase (day 8-13) Early luteal phase (day 14-22) Late luteal phase (day 23-28)**	**10 women who have regular menstrual cycle, and 10 women who use hormonal contraceptive.**	**PTA**	**Day counting**	**There is a significant effect of hormonal fluctuations and hearing thresholds across the menstrual cycle. The hearing threshold of women who don't use contraceptive varied significantly through the cycle with mean variation of 4.09 dB HL. Men hearing threshold did not varied between the sessions. For women who did not use contraceptives, the lowest threshold was observed in the late follicular phase.**
Emami et al., 2018	Case-control study	Women (20)	19-30	Normal hearing	No control	Two sessions: Late follicular phase (day 13). Early luteal phase (day 22).	Regular cycle (28 days)	PTA Tympanometry DPOAEs ABR	Not mentioned	It was reported that there are individual differences in the effect of female sex hormones on hearing. As the auditory function seems to be sensitive in some women to hormonal changes. During the early luteal phase, high level of progesterone caused a reduction in hearing (worse hearing at 250 Hz), increased DPOAEs amplitude, delayed ABR interpeak latencies). Better hearing sensitivity in follicular phase.
Carneiro et al., 2019	Cohort, longitudinal, blinded	Women (9) Men (11)	25 (15)	Normal hearing	Age-matched men	Two sessions for both groups For women: Late follicular phase (day 11-13) Late luteal phase (day 23-26)	Regular menstrual cycles (27 to 32 days)	Dichotic testing: SSW, DD, and CV	Blood sample	In late follicular phase, better responses in women and in the right ear. Estradiol improved dichotic listening in women during higher level of E2 in the menstrual cycle. No significant changes in men performances.

(Continued)

TABLE 2 (Continued)

References	Study design	Sample size	Age mean (SD)	Hearing level	Control group	Number of sessions	Experimental group description	Outcome measures		Findings
								auditory	hormones	
Karaer and Gorkem, 2020	Case-control, cross-sectional	Women (90; 30 premature ovarian failure, 30 normal, 30 menopausal)	POF: 32.5 (1.06), menopause 54.4 (1.1), Control 28.4 (1.06)	POF, control: normal hearing. older women: normal hearing, mild HL at 6-8 Hz	Normal pre-menopausal women	Each group was tested in one session: For the POF tested after at least 6 months of the diagnosis (the presence of amenorrhea and a serum FSH level greater than 40 mIU/mL on two occasions at least 1 month apart). For the menopausal women (amenorrhea for at least 1 year). For the control group (normal regular menstrual cycle), no mention of the cycle phase or day of the testing.	Hearing function is impaired in women with premature ovarian failure	PTA TEOAEs DPOAEs	Blood sample	In the POF group, there was a significant decline in the 6 kHz DPOAEs compared with the pre-menopausal group. The menopausal group showed worse hearing which reported to be due to aging and reduced level of estradiol. There was a negative effect of reduced estradiol levels on OHC function, in the POF group. Normal estradiol levels may promote healthy OHC function.

3. Results

3.1. Sex differences between pre-menopausal women and age-matched men

3.1.1. Peripheral auditory function

The findings suggested an overall sex difference for measures of peripheral auditory function, particularly PTA. Specifically, with nine studies out of eleven showed a significant sex-specific differences in the audiometric measures in favor of women with PTA thresholds in adults (between 18 and 49 years) being better in women than men, mainly at higher frequencies (Dreisbach et al., 2007; Sharashenidze et al., 2008; Kim et al., 2010).

Although sex differences were also evident in DPOAEs recordings, they have been suggested to be related to the anatomical differences (i.e., due to the differences in the cochlea length) rather than the biological sex differences (Bowman et al., 2000; Dreisbach et al., 2007; Boothalingam et al., 2018). However, all other types of OAEs suggested significant sex differences in the function of the cochlea [i.e., outer hair cells (OHCs)] independent of its length. Women and right ears were reported to have stronger SOAEs and larger TEOAEs, and contralateral TEOAEs amplitudes (Ismail and Thornton, 2003; Snihur and Hampson, 2011; Stuart and Kerls, 2018).

3.1.2. Central auditory system

ABR recording at suprathreshold levels showed sex differences between pre-menopausal women and men with women showing better responses, i.e., shorter latencies and larger amplitudes (Zakaria et al., 2019). In addition, speech-ABRs showed larger amplitudes and shorter latencies of wave V and A in women compared to men (Jalaei et al., 2017). 40-Hz ASSRs were reported to be better in women than men, however, this was only reported for left-handed but not for right-handed participants (Melynyte et al., 2018).

Evoked potential recordings showed sex differences. Sex differences in ABRs were found only at suprathreshold levels, where women had better responses than men (i.e., larger amplitude and shorter latencies). These differences remained even when controlling for the differences in head sizes (Zakaria et al., 2019) and were reported to be not related to the differences in head size of participants. Don et al. (1993) reported the same findings. In addition, speech-ABRs were also found to be better in women with larger amplitude of waves as compared to men. However, the latencies of speech-ABRs were reported to be related to anatomical differences of the head diameter (Jalaei et al., 2017). The ABR recordings of menopausal women showed prolonged conduction time which was reported to be the driven by changes in female sex hormones levels.

3.2. Female sex hormone fluctuation in pre-menopausal women

3.2.1. Peripheral auditory function

Hearing sensitivity was reported to be improved during the late follicular phase of the cycle (i.e., mainly during higher

TABLE 3 Summary of the characteristics and results of studies on "the effect of menopause on hearing function."

References	Study design	Sample size	Mean age (SD)	Hearing level	Control group	Number of sessions	Experimental group description	Outcome measures		Findings
								auditory	hormones	
Tandon et al., 2001	Observational study-cross-sectional	Women (22)	Post-menopausal between 50 and 70	Normal hearing	No control	One session	Post-menopausal	ABR	Not mentioned	Post-menopausal women had longer conduction time due to hormonal changes resulting from menopause. Significant increase in wave I, III, V latencies and the interpeak latency between I-III, I-V, and III-V in post-menopausal women
Hederstierna et al., 2010	Observational study - longitudinal	Women (104)	51.2 (1.5)	baseline: normal hearing	No control	Tested twice with an average interval of 7.5 years	Post-menopausal	PTA	Not mentioned	It was reported that a rapid decline of hearing levels in healthy women after the start of menopause, which appears to act as a trigger of age-related hearing loss in women. This decline was noticed to start in the left ear.
Svedbrant et al., 2015	Observational study - longitudinal	Women (100)	49.3 (1.6)	Baseline: normal hearing	No control	2, 7, 10 years follow-up	Post-menopausal	PTA	Blood sample	The hearing level declined rapidly in the peri-menopausal group at 1-3 kHz for both ears, and a rapid decline of hearing was seen after menopause. However, no significant correlation between hormonal levels and hearing levels for this age group.
Trott et al., 2019	A prospective, group comparison study	14 Peri-post-menopausal women	54 years	Normal hearing	Pre-menopausal women	One session	Peri-menopausal women (Having irregular cycles between three and 11 months) Post-menopausal women (Having at least 1 year of amenorrhea)	Dichotic Digit (DD) testing Speech in noise (LiSN-S/SPIN-R) ABR- MLR	Not mentioned	Non-significant differences in DD, SPIN-R tests or MLR between groups. Significant differences in LiSN-S between groups, pre- and post-menopausal women have poor SRT. Significant ABR differences, pre- and post-menopausal women have longer wave V latencies with a higher stimulus rate.
Arora et al., 2021	Cross-sectional	Pre-menopausal women (90) Post-menopausal women (100)	18.6 (0.73) 59.8 (5.84)	Both groups have normal hearing	Pre-menopausal women were tested during follicular phase	One session	Post-menopausal women (at least 1 year of amenorrhea)	ABR	Not mentioned	ABR waves latencies increased in post-menopausal women which show subtle degenerative changes in hearing that start appearing in the central auditory pathway after menopause and probably caused by estradiol decline. As estradiol levels influences the sensory transmission in the auditory pathway.

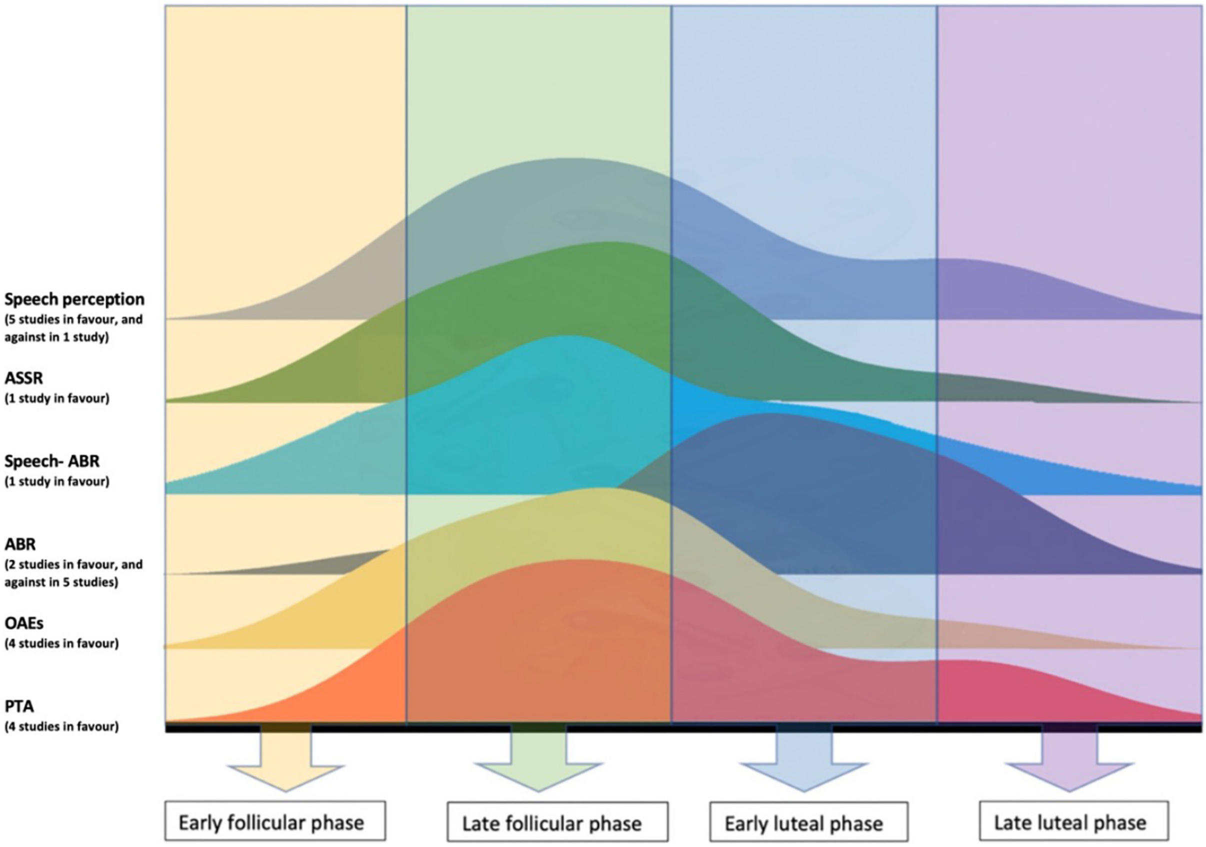


FIGURE 3
Illustration of the fluctuation of the audiological performances across the menstrual cycle, the peaks represent better performance.

levels of estradiol). PTA thresholds were found to be decreased (i.e., better hearing sensitivity) during the late follicular phase (Adriztina et al., 2016; Souza et al., 2017; Emami et al., 2018; Karaer and Gorkem, 2020).

Spontaneous Otoacoustic Emissions, TOAEs and DPOAEs were better during late follicular phase (Al-Mana et al., 2010; Adriztina et al., 2016; Emami et al., 2018; Karaer and Gorkem, 2020). On the other hand, no significant effect of progesterone on OAEs was reported (Al-Mana et al., 2010).

A negative correlation between MOC suppression and estradiol was reported, and no significant effect of progesterone on MOC suppression (Al-Mana et al., 2010).

3.2.2. Central auditory system

Reports regarding the effect of estradiol and progesterone on ABR wave latencies were inconsistent. Some studies reported increased latencies (i.e., longer transmission time) of ABR I-V waves (i.e., worsening) during the late follicular phase and shorter latencies (i.e., improvement) were reported to be during luteal phase (i.e., during higher levels of progesterone) (Al-Mana et al., 2010; Mann et al., 2012; Upadhayay et al., 2014; Hu and Lau, 2017; Emami et al., 2018). Other studies reported decreased ABR latencies (i.e., improvement) during follicular phase (Serra et al., 2003; Batta et al., 2017). There is an increase in the amplitude of speech-ABR waves during late follicular phase, but no changes in the latencies were reported (Liu et al., 2017). 40-Hz ASSR have been reported to improve during higher levels of estradiol in one study (Griskova-Bulanova et al., 2014).

Long latency auditory evoked responses (LLEAPs) recording was found to fluctuate in women with normal ovulatory cycle, however, there was no fluctuation of LLEAPs recording in women with anovulatory cycle (who use hormonal contraceptives) (Yadav et al., 2003). In addition, better ERPs were reported to occur during luteal phase only, i.e., when the level of progesterone increases (Walpurger et al., 2004). The following figure illustrates the fluctuation of the audiological tests results during the menstrual cycle as reported by the included studies.

The performance of women in speech audiometry fluctuated through the menstrual cycle. Five out of six studies reported better performance in speech perception during high levels of estradiol (Cowell et al., 2011; Hjelmervik et al., 2012; Hodgetts et al., 2015; Hu and Lau, 2017; Carneiro et al., 2019), whereas the sixth study, (Wadnerkar et al., 2008) reported no significant effect of estradiol in dichotic listening during the follicular phase, and no differences in response number between women and men. Wadnerkar et al. (2008) findings could not reflect the true effect of estradiol in hearing for two reasons. First, the day of the menstrual cycle was self-reported by participants so the level of female sex hormones can only be surmised. Using a self-reported measure to investigate the level of hormones in the body is known not to be accurate. Another explanation to this finding, Wadnerkar et al. (2008) tested women in two sessions: one session

TABLE 4 Summary of the studies' findings.

	Static sex differences		Cyclical changes/hormonal fluctuation	
	Pre-menopausal vs. age-matched men	Post-menopausal vs. age-matched men	Pre-menopausal women	Post-menopausal women
PTA	Women have better performance	Post-menopausal women tend to have steeper decreased hearing sensitivity than men	Better performance during late follicular phase	Fast and rapid decline in hearing in high frequency after the start of menopause.
TEOAEs SOAEs MLSOAEs CEOAEs	Women have better performance		Better performance during late follicular phase	
ABR	Women have better performance		Better performance during early luteal phase reported by most of the studies, however, there was a conflicted result, as it was reported better performance was during late follicular phase.	Longer waves latencies in women between 50–70 years old.
Speech perception	Women have better performance		Better performance during late follicular phase	Poor performance after menopause.
DPOAEs	No significant sex differences in the recordings. The differences are related to the anatomical differences (i.e., the length of the cochlea)		Better performance in late follicular phase Worse recording in the premature ovarian failure (POF) group	
Speech-ABR	Women have better performance		Worse performance during late follicular phase	
40 Hz ASSR	Left-handed women had better performance than left-handed men. No significant differences between right-handed women and men.		Better performance during late follicular phase	
Dichotic testing			Better right ear responses during late follicular phase	Poor responses after menopause.

in the early follicular phase (day 2–5) which was during low estradiol and progesterone levels, and another session which fell between two phases, the early and late luteal phase (day 18–25). The second session reported by Wadnerkar et al. (2008) to be in the follicular phase and during high levels of estradiol and progesterone. However, since all participants were reported of having normal average menstrual cycle (around 28 days) then this session was undertaken in luteal phase and not the follicular phase. **Figure 3** illustrates the fluctuation of the audiological performances across the menstrual cycle phases, where the peaks represent better performance.

3.3. Auditory changes in post-menopausal women

A significant rapid reduction in hearing sensitivity after menopause has been reported, particularly at 1 kHz (Hederstierna et al., 2010) and 3 kHz (Svedbrant et al., 2015). Whether there is an ear asymmetry to this decline is unclear as one study found it more pronounced in the right ear (Svedbrant et al., 2015), the other in the left ear (Hederstierna et al., 2010). In addition to peripheral hearing sensitivity, ABR waves latencies were also increased after the start of menopause (Tandon et al., 2001; Trott et al., 2019; Arora et al., 2021). Significant differences in speech reception in noise, as poor performance was found in pre- and post-menopausal women with normal PTA thresholds, suggesting some central hearing loss. The findings of the studies are summarized in **Table 4**.

4. Quality of evidence

The quality and risk of bias of the included studies was assessed using Newcastle-Ottawa Scale (NOS). Only four studies were of good quality (Hjelmervik et al., 2012; Melynyte et al., 2018; Carneiro et al., 2019; Zakaria et al., 2019). 27 studies were of fair quality (high risk) (Bowman et al., 2000; Ismail and Thornton, 2003; Serra et al., 2003; Yadav et al., 2003; Dreisbach et al., 2007; Sharashenidze et al., 2008; Wadnerkar et al., 2008; Al-Mana et al., 2010; Hederstierna et al., 2010; Kim et al., 2010; Cowell et al., 2011; Snihur and Hampson, 2011; Svedbrant et al., 2015; Adriztina et al., 2016; Hu and Lau, 2017; Jalaei et al., 2017; Boothalingam et al., 2018; Trott et al., 2019; Zakaria et al., 2019; Karaer and Gorkem, 2020; Arora et al., 2021). Four studies were of poor quality (very high risk) (Tandon et al., 2001; Griskova-Bulanova et al., 2014; Hodgetts et al., 2015; Emami et al., 2018).

The main concern was the method of assessment for hormone levels, as few studies used objective tests such as blood assays and saliva samples. Another factor that affected the quality of the studies which examined the effect of female hormone fluctuation on hearing, was the number of sessions. Only three studies were considered to have a "good" number of sessions for the studied outcomes to occur, as they tested participants in three or four sessions across one menstrual cycle. Finally, most studies included in the review did not have control group. The quality of the studies is summarized in **Table 5**.

TABLE 5 Quality and risk of bias assessment (Newcastle–Ottawa Scale) criteria.

	Selection			Comparability			Outcome		Total quality score	
	Representativeness of the exposed cohort	Selection of the non-exposed cohort	Ascertainment of exposure	Demonstration that outcome of interest was not present at start of study	Comparability of cohorts based on the design or analysis	Assessment of outcome	Was follow-up long enough for outcomes to occur	Adequacy of follow up of cohorts		
References										
Bowman et al., 2000		*		*	*	*			4	Fair quality
Tandon et al., 2001	*					*			2	Poor quality
Ismail and Thornton, 2003		*		*	*	*			4	Fair quality
Dreisbach et al., 2007		*		*	*	*	*		5	Fair quality
Sharashenidze et al., 2008	*	*		*		*	*		5	Fair quality
Kim et al., 2010	*	*		*	*	*			5	Fair quality
Snihur and Hampson, 2011		*		*	*	*			4	Fair quality
Jalaei et al., 2017		*		*	*	*	*		6	Fair quality
Boothalingam et al., 2018	*	*		*	*	*			5	Fair quality
Melynyte et al., 2018	*	*	*	*	*	*	*		7	Good quality
Stuart and Kerls, 2018				*	*	*	*		4	Fair quality
Zakaria et al., 2019	*	*	*	*	*	*	*		7	Good quality
Serra et al., 2003	*		*			*	*	*	5	Fair quality
Yadav et al., 2003		*		*	*	*	*	*	6	Fair quality
Walpurger et al., 2004			*	*	*	*	*		5	Fair quality
Wadnerkar et al., 2008	*	*		*	*	*		*	6	Fair quality
Al-Mana et al., 2010	*		*	*		*	*	*	6	Fair quality

(Continued)

TABLE 5 (Continued)

	Selection			Comparability		Outcome		Total quality score	
	Representativeness of the exposed cohort	Selection of the non-exposed cohort	Ascertainment of exposure	Demonstration that outcome of interest was not present at start of study	Comparability of cohorts based on the design or analysis	Assessment of outcome	Was follow-up long enough for outcomes to occur	Adequacy of follow up of cohorts		
References										
Hederstierna et al., 2010	*			*		*	*	*	**5**	Fair quality
Cowell et al., 2011	*		*	*		*		*	**5**	Fair quality
Hjelmervik et al., 2012	*	*	*	*	*	*	*		**7**	Good quality
Mann et al., 2012				*	*	*	*		**4**	Fair quality
Griskova-Bulanova et al., 2014			*	*		*			**3**	Poor quality
Upadhayay et al., 2014				*	*	*	*		**4**	Fair quality
Hodgetts et al., 2015			*	*		*			**3**	Poor quality
Svedbrant et al., 2015	*		*	*		*	*		**5**	Fair quality
Adriztina et al., 2016	*			*		*	*	*	**5**	Fair quality
Batta et al., 2017				*	*	*	*	*	**5**	Fair quality
Hu and Lau, 2017				*		*	*	*	**4**	Fair quality
Liu et al., 2017		*	*	*	*	*	*		**6**	Fair quality
Souza et al., 2017		*		*	*	*	*		**5**	Fair quality
Emami et al., 2018				*		*		*	**3**	Poor quality
Carneiro et al., 2019	*	*	*	*	*	*		*	**7**	Good quality
Trott et al., 2019	*	*		*	*			*	**5**	Fair quality
Karaer and Gorkem, 2020		*		*	*	*			**4**	Fair quality
Arora et al., 2021	*	*		*	*			*	**5**	Fair quality

NOS has a total maximum score of 9: Maximum scores 4 in Selection, 2 in Comparability, 3 in Outcome. Studies score from 7–9 have good quality (high quality), 4–6 have fair quality (high risk), and 0–3 have poor quality (very high risk).
The symbol (*) means the point earned in each category.

5. Discussion

The aim of the systematic review was to evaluate the current evidence of the differences in auditory function between women and men. In addition, the aim was to review the available literature of the effect of the female sex hormones (i.e., estradiol and progesterone) on fluctuating auditory function in women (i.e., during the menstrual cycle and after menopause).

Eleven studies investigated sex-specific differences in the peripheral and central auditory pathways. They reported that women's hearing sensitivity was better compared to age-matched men's, especially at higher frequencies (Kim et al., 2010). Participants age ranged from 15 to 83 years. The mean age of men was 46 years, and the mean age of women was 47 years (which might be considered before menopause). Most women were 30–39 years old (n = 242) and 40–49 years old (n = 313). The total pooled sample size of this review was large (n = 1,116). The consistent finding of the review of better hearing sensitivity of pre-menopausal women compared to men is in agreement with a previous study (n = 50,000) that pre-menopausal women have better hearing sensitivity than men (Chung et al., 1983), in particular at higher frequencies (approximately 2–3.5 dB differences at frequencies above 2,000 Hz). Another consistent finding across studies was that hearing sensitivity of pre-menopausal women fluctuates across the menstrual cycle, while men tend to show stable hearing sensitivity. In terms of cyclical changes, PTA thresholds were found to be lowest (i.e., better hearing sensitivity) during the late follicular phase compared to other phases of the cycle (Adriztina et al., 2016; Souza et al., 2017; Emami et al., 2018; Karaer and Gorkem, 2020).

It can be argued that the reduction in hearing sensitivity in older women could be due to normal aging, noise exposure, and ototoxicity. However, the changes in hearing were found to be triggered by the onset of menopause (Hederstierna et al., 2009). And similar changes have also been seen in women with premature ovarian failure (POF). In particular, POF and post-menopausal women groups experienced reduced hearing function compared to normal pre-menopausal women (Karaer and Gorkem, 2020). In addition, while ear asymmetries in hearing loss are inconsistent, better hearing in the right ear could be explained anatomically by the number of estradiol receptors in the inner ear. McFadden (1993) reported that the right inner ear is denser in estradiol receptors than in the left ear. These receptors facilitate the effect of estradiol in the inner ear cells, which may enhance the transition of neural signals from the right ear. Once the level of estradiol is reduced in POF or post-menopause, the reduction in that ear may be particularly noticeable.

Like PTA thresholds, SOAEs and TOAEs were reported to be stronger in women (Ismail and Thornton, 2003; Snihur and Hampson, 2011). The function of OHCs might be better in women when compared to men. This might indicate a fluctuation in the inner ear function because of changes in the female sex hormones.

However, several researchers have suggested that these differences may be due to the anatomical differences in the cochlea's length rather than related to the biological (Bowman et al., 2000; Dreisbach et al., 2007; Boothalingam et al., 2018). In summary, the results suggest that DPOAEs might not be a useful measure to detect sex differences in the auditory function.

The effect of hormones on central auditory function is less clear. While some consistent evidence exists that pre-menopausal women have better overall central auditory functioning, it is less clear whether there are consistent changes in central auditory function across the menstrual cycle. These effects are exemplified in ABR measures. Specifically, fairly robust evidence exists for sex differences in ABRs at suprathreshold levels, with women generally having better responses than men. When inconsistencies between ABR results were reported, particularly in latencies, a possible explanation may be the variation in session numbers and the use of objective measures for female hormones. The possible contribution of estradiol and progesterone in the central auditory pathways may remain unclear, and whether estradiol or progesterone can improve conduction of auditory neural signals. However, the effect of reduced levels of estradiol in post-menopausal women were found to affect first the central auditory pathway (Trott et al., 2019; Arora et al., 2021). It was found that post-menopausal women with normal hearing sensitivity have longer ABR waves latencies (Hwang et al., 2008; Trott et al., 2019; Arora et al., 2021).

For the studies which included men as control, only fluctuation in hearing was reported in women. Therefore, men may have a stable hearing sensitivity. In other words, due to stable levels of female hormones in men, a stable hearing function was noticed. However, this was reported in only three papers (Hjelmervik et al., 2012; Souza et al., 2017; Carneiro et al., 2019) as both sexes were tested in 3–4 sessions across the cycle, and one study tested both sexes in one session (Liu et al., 2017) and one study tested men in one session only and women in two sessions (Wadnerkar et al., 2008). The variation in the design of papers studied the effect of female sex hormones may cause uncertainty regarding the interpretation of the role of these hormones.

This is the first systematic review that has attempted to address differences in auditory function between the sexes and the possible effect of female sex hormone fluctuation on hearing function. The conclusion of this review is drawn from thirty-three studies. The lack of "good" quality studies makes it challenging to understand the effect of female hormones on hearing in detail. The review highlights the need for objective measures to assess the hormone level at the time of testing. In addition, participants need to be tested in multiple, ideally four or more, sessions throughout the menstrual cycle to detect the effect of hormone changes on hearing, so that errors in test timing can be avoided.

Most of the studies were not controlled, and only three studies included male participants as a control group. All studies conducted in menopausal women did not use any control groups.

In addition, in order to improve objectivity of measures researcher could consider using a blind study design and objective tests such as blood or saliva samples to measure hormones levels. None of the studies included in this review stated the day of the cycle when women were tested. Accurately measuring and reporting this information may help to disambiguate some of the currently inconsistent results.

No studies on the possible effects of hormones overall or fluctuation of estradiol and progesterone on tinnitus or vestibular dysfunction existed highlighting the severe lack of studies on this topic.

In conclusion, there are significant sex differences in peripheral auditory function, particularly PTA threshold, SOAEs, TEOAEs, between pre-menopausal women and age-matched men. In addition, a possible effect of estradiol on peripheral auditory function across the menstrual cycle was reported in most of the included papers. In contrast, the effect of estradiol and progesterone in the central auditory system remains unclear. Whether this difference in results between peripheral and central auditory function reflects a true difference in function or a difference in assessment is currently unclear. PTA is the main tool used in audiology clinics and research, hence more evidence, and importantly more consistent evidence, can accumulate. Tests that assess speech reception in background noise are less frequently used both in research and in the clinic despite their greater usefulness to assess aspects of hearing that are important for everyday listening. This can be an important tool to assess higher regions of the auditory pathway, including cognition. A more frequent use would allow us to build up a more detailed picture of the effect of sex hormones overall and their effect across the menstrual cycle. Finally, it was noticeable how much outcome measures differed between studies, and that the majority of studies did not use an objective test to measure hormones levels. It is recommended for the future studies to include consistent outcome measure which may include audiometric tests such as PTA (including extended high frequencies) and speech audiometry (e.g., SiN).

5.1. Deviation from the published protocol

The protocol was restricted to studies with control groups. However, this restriction excluded many studies that investigated the fluctuation of female hormones and changes in hearing sensitivity. These studies used objective hormonal tests and a greater number of sessions. Therefore, papers without control groups were included in the review, but their qualities were affected by that.

Author contributions

NA wrote the manuscript with input from all authors. AH, KM, and KK edited the manuscript. All authors discussed the results and contributed to the final version of the manuscript.

Funding

This research was funded by Taibah University and supported by the NIHR Manchester Biomedical Research Centre (NIHR203308).

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2023.1077409/full#supplementary-material

References

Abdala, C., and Visser-Dumont, L. (2001). Distortion product otoacoustic emissions: a tool for hearing assessment and scientific study. *Volta Rev.* 103, 281–302.

Adriztina, I., Adnan, A., Adenin, I., Haryuna, S., and Sarumpaet, S. (2016). Influence of hormonal changes on audiologic examination in normal ovarian cycle females: an analytic study. *Int. Arch. Otorhinolaryngol.* 20, 294–299. doi: 10.1055/s-0035-1566305

Al-Mana, D., Ceranic, B., Djahanbakhch, O., and Luxon, L. (2008). Hormones and the auditory system: a review of physiology and pathophysiology. *Neuroscience* 153, 881–900.

Al-Mana, D., Ceranic, B., Djahanbakhch, O., and Luxon, L. (2010). Alteration in auditory function during the ovarian cycle. *Hear Res.* 268, 114–122. doi: 10.1016/j.heares.2010.05.007

Anias, C. R., Lima, M. A. M. T., and Kos, A. O. A. (2004). Evaluation of the influence of age in auditory brainstem response. *Brazil. J. Otorhinolaryngol.* 70, 84–89.

Arch, V. S., and Narins, P. M. (2009). Sexual hearing: the influence of sex hormones on acoustic communication in frogs. *Hear. Res.* 252, 15–20. doi: 10.1016/j.heares.2009.01.001

Arora, S., Mittal, S., Gupta, S., Loona, S., Singh, K., and Mehra, K. (2021). Comparative analysis of brainstem auditory evoked potential patterns between menstruating and menopausal north Indian females. *Indian J. Endocrinol. Metab.* 25, 438–442. doi: 10.4103/ijem.ijem_368_21

Ashok Murthy, V., and Krishna, K. (2013). Hearing loss in pregnancy. *Indian J. Otolaryngol. Head Neck Surg.* 65, 1–2. doi: 10.1007/s12070-011-0329-0

Batta, M., Dhir, S., Kumar, A., and Singh, K. (2017). Effect of different phases of menstrual cycle on brainstem auditory evoked response. *Int. J. Appl. Basic Med. Res.* 7, 44–47. doi: 10.4103/2229-516X.198522

Bonnard, Å, Bark, R., and Hederstierna, C. (2018). Clinical update on sensorineural hearing loss in Turner syndrome and the X-chromosome. *Am. J. Med. Genet. C Semin. Med. Genet.* 181C, 67–73. doi: 10.1002/ajmg.c.31673

Bonnard, Å, Hederstierna, C., Bark, R., and Hultcrantz, M. (2017). Audiometric features in young adults with Turner syndrome. *Int. J. Audiol.* 56, 650–656. doi: 10.1080/14992027.2017.1314559

Bonnard, A., Sahlin, L., Hultcrantz, M., and Simonoska, R. (2013). No direct nuclear effect of progesterone in the inner ear: other possible pathways. *Acta Otolaryngol.* 133, 1250–1257. doi: 10.3109/00016489.2013.825377

Boothalingam, S., Klyn, N. A. M., Stiepan, S. M., Wilson, U. S., Lee, J., Siegel, J. H., et al. (2018). "Revisiting gender, race, and ear differences in peripheral auditory function," in *Proceedings of the 13th Mechanics of Hearing Workshop*, (St. Catharines, ON: AIP Publishing).

Bowman, D., Brown, D., and Kimberley, B. (2000). An examination of gender differences in DPOAE phase delay measurements in normal-hearing human adults. *Hear. Res.* 142, 1–11. doi: 10.1016/s0378-5955(99)00212-9

Burns, E., Arehart, K., and Campbell, S. (1992). Prevalence of spontaneous otoacoustic emissions in neonates. *J. Acoust. Soc. Am.* 91, 1571–1575. doi: 10.1121/1.402438

Canlon, B., and Frisina, R. D. (2009). Sex hormones in the auditory system. *Hear. Res.* 252, 1–2.

Carneiro, C., Almeida, A., Ribas, A., Kluk-De Kort, K., Lima, D., Araújo, A., et al. (2019). Hormones and auditory perception: study of dichotic listening in women during the menstrual cycle. *Int. Arch. Otorhinolaryngol.* 23, 70–76. doi: 10.1055/s-0038-1626701

Caruso, S., Cianci, A., Grasso, D., Agnello, C., Galvani, F., Maiolino, L., et al. (2000). Auditory brainstem response in postmenopausal women treated with hormone replacement therapy: a pilot study. *Menopause* 7, 178–183. doi: 10.1097/00042192-200007030-00008

Caruso, S., Maiolino, L., Agnello, C., Garozzo, A., Di Mari, L., and Serra, A. (2003). Effects of patch or gel estrogen therapies on auditory brainstem response in surgically postmenopausal women: a prospective, randomized study. *Fertil. Steril.* 79, 556–561. doi: 10.1016/s0015-0282(02)04763-5

Charitidi, K., and Canlon, B. (2010). Estrogen receptors in the central auditory system of male and female mice. *Neuroscience* 165, 923–933.

Charitidi, K., Frisina, R., Vasilyeva, O., Zhu, X., and Canlon, B. (2010). Expression patterns of estrogen receptors in the central auditory system change in prepubertal and aged mice. *Neuroscience* 170, 1270–1281. doi: 10.1016/j.neuroscience.2010.08.010

Charitidi, K., Meltser, I., Tahera, Y., and Canlon, B. (2009). Functional responses of estrogen receptors in the male and female auditory system. *Hear. Res.* 252, 71–78.

Chlebowski, R., Anderson, G., Aragaki, A., Manson, J., Stefanick, M., Pan, K., et al. (2020). Association of menopausal hormone therapy with breast cancer incidence and mortality during long-term follow-up of the women's health initiative randomized clinical trials. *JAMA* 324, 369–380. doi: 10.1001/jama.2020.9482

Chung, D. Y., Mason, K., Gannon, R. P., and Willson, G. N. (1983). The ear effect as a function of age and hearing loss. *J. Acoust. Soc. Am.* 73, 1277–1282.

Clarivate Analytics (2018). *EndNote X9. 64 bit*. Philadelphia, PA: Clarivate Analytics.

Clayton, J., and Collins, F. (2014). Policy: NIH to balance sex in cell and animal studies. *Nature* 509, 282–283. doi: 10.1038/509282a

Contoreggi, N., Mazid, S., Goldstein, L., Park, J., Ovalles, A., Waters, E., et al. (2021). Sex and age influence gonadal steroid hormone receptor distributions relative to estrogen receptor β-containing neurons in the mouse hypothalamic paraventricular nucleus. *J. Comp. Neurol.* 529, 2283–2310. doi: 10.1002/cne.25093

Corso, J. F. (1968). Age and sex differences in pure-tone thresholds: a survey of hearing levels from 18 to 65 years. *Am. Found. Blind Res. Bull.* 17, 141–172. doi: 10.1001/archotol.1963.00750010399008

Corwin, J., Bullock, T., and Schweitzer, J. (1982). The auditory brain stem response in five vertebrate classes. *Electroencephalogr. Clin. Neurophysiol.* 54, 629–641. doi: 10.1016/0013-4694(82)90117-1

Cowell, P., Ledger, W., Wadnerkar, M., Skilling, F., and Whiteside, S. (2011). Hormones and dichotic listening: evidence from the study of menstrual cycle effects. *Brain Cogn.* 76, 256–262. doi: 10.1016/j.bandc.2011.03.010

Davis, A. (1995). *Hearing in adults*, 1st Edn. London: Whurr.

Dehan, C., and Jerger, J. (1990). Analysis of gender differences in the auditory brain-stem response. *Laryngoscope* 100, 18–24.

Delhez, A., Lefebvre, P., Péqueux, C., Malgrange, B., and Delacroix, L. (2020). Auditory function and dysfunction: estrogen makes a difference. *Cell Mol. Life Sci.* 77, 619–635. doi: 10.1007/s00018-019-03295-y

Di Renzo, G., Mattei, A., Gojnic, M., and Gerli, S. (2005). Progesterone and pregnancy. *Curr. Opin. Obstet. Gynecol.* 17, 598–600. doi: 10.1097/01.gco.0000191899.84567.4d

Disney, A., and Calford, M. (2001). Neurosteroids mediate habituation and tonic inhibition in the auditory midbrain. *J. Neurophysiol.* 86, 1052–1056. doi: 10.1152/jn.2001.86.2.1052

Don, M., Ponton, C., Eggermont, J., and Masuda, A. (1993). Gender differences in cochlear response time: an explanation for gender amplitude differences in the unmasked auditory brain-stem response. *J. Acoust. Soc. Am.* 94, 2135–2148. doi: 10.1121/1.407485

Dreisbach, L., Kramer, S., Cobos, S., and Cowart, K. (2007). Racial and gender effects on pure-tone thresholds and distortion-product otoacoustic emissions (DPOAEs) in normal-hearing young adults. *Int. J. Audiol.* 46, 419–426. doi: 10.1080/14992020701355074

Emami, S., Gohari, N., Ramezani, H., and Borzouei, M. (2018). Hearing performance in the follicular-luteal phase of the menstrual cycle. *Int. J. Otolaryngol.* 2018, 7276359. doi: 10.1155/2018/7276359

Farrar, D., Neill, J., Scally, A., Tuffnell, D., and Marshall, K. (2015). Is objective and accurate cognitive assessment across the menstrual cycle possible? A feasibility study. *SAGE Open Med.* 3:2050312114565198. doi: 10.1177/2050312114565198

Folsom, R. C., Burns, E. M., Zettner, E. M., and Morrison, R. B. (1994). Transient-evoked otoacoustic emission suppression tuning curves in human adults and infants. *J. Acoust. Soc. Am.* 95, 2844–2844. doi: 10.1121/1.2977737

Frisina, R., (2012). Hormones and hearing: too much or too little of a good thing can be ototoxic. *Thieme Semin. Hear.* 33, 231–241. doi: 10.1016/j.heares.2020.108093

Gelfand, S. (2004). *An introduction to psychological and physiological acoustics*, 4 Edn. New York, NY: Marcel Dekker.

Gold, T. (1948). Hearing. II. The physical basis of the action of the cochlea. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* 135, 492–498.

Grabham, N. J., Swabey, M. A., Chambers, P., Lutman, M. E., White, N. M., and Chad, J. E. (2013). An evaluation of otoacoustic emissions as a biometric. *IEEE Trans. Inf. Forens. Secur.* 8, 174–183.

Grinn, S., Wiseman, K., Baker, J., and Le Prell, C. (2017). Hidden hearing loss? No effect of common recreational noise exposure on cochlear nerve response amplitude in humans. *Front. Neurosci.* 11:465. doi: 10.3389/fnins.2017.00465

Griskova-Bulanova, I., Griksiene, R., Korostenskaja, M., and Ruksenas, O. (2014). 40 Hz auditory steady-state response in females: when is it better to entrain? *Acta Neurobiol. Exp.* 74, 91–97.

Guimaraes, P., Frisina, S., Mapes, F., Tadros, S., Frisina, D., and Frisina, R. (2006). Progestin negatively affects hearing in aged women. *Proc. Natl. Acad. Sci. U.S.A.* 103, 14246–14249. doi: 10.1073/pnas.0606891103

Hawkins, S., and Matzuk, M. (2008). The menstrual cycle. *Ann. N. Y. Acad. Sci.* 1135, 10–18.

He, Z.-Y., and Ren, D. (2018). Sex hormones and inner ear. *IntechOpen* 14, 329–346.

Hederstierna, C., Hultcrantz, M., and Rosenhall, U. (2009). Estrogen and hearing from a clinical point of view; characteristics of auditory function in women with turner syndrome. *Hear. Res.* 252, 3–8. doi: 10.1016/j.heares.2008.11.006

Hederstierna, C., Hultcrantz, M., Collins, A., and Rosenhall, U. (2010). The menopause triggers hearing decline in healthy women. *Hear. Res.* 259, 31–35. doi: 10.1016/j.heares.2009.09.009

Hjelmervik, H., Westerhausen, R., Osnes, B., Endresen, C., Hugdahl, K., Hausmann, M., et al. (2012). Language lateralization and cognitive control across the menstrual cycle assessed with a dichotic-listening paradigm. *Psychoneuroendocrinology* 37, 1866–1875. doi: 10.1016/j.psyneuen.2012.03.021

Hodgetts, S., Weis, S., and Hausmann, M. (2015). Sex hormones affect language lateralisation but not cognitive control in normally cycling women. *Horm. Behav.* 74, 194–200. doi: 10.1016/j.yhbeh.2015.06.019

Hu, X., and Lau, C. (2017). Central conduction time in auditory brainstem response and ear advantage in dichotic listening across menstrual cycle. *PLoS One* 12:e0187672. doi: 10.1371/journal.pone.0187672

Hultcrantz, M., Simonoska, R., and Stenberg, A. (2006). Estrogen and hearing: a summary of recent investigations. *Acta Otolaryngol.* 126, 10–14. doi: 10.1080/00016480510038617

Hwang, J., Chao, J., Ho, H., and Hsiao, S. (2008). Effects of sex, age and hearing asymmetry on the interaural differences of auditory brainstem responses. *Audiol. Neurootol.* 13, 29–33. doi: 10.1159/000107468

Ismail, H., and Thornton, A. (2003). The interaction between ear and sex differences and stimulus rate. *Hear. Res.* 179, 97–103. doi: 10.1016/s0378-5955(03)00099-6

Jalaei, B., Zakaria, M., Mohd Azmi, M., Nik Othman, N., and Sidek, D. (2017). Gender disparities in speech-evoked auditory brainstem response in healthy adults. *Ann. Otol. Rhinol. Laryngol.* 126, 290–295. doi: 10.1177/0003489417690169

Jerger, J., Silman, S., Lew, H. L., and Chmiel, R. (1993). Case studies in binaural interference: converging evidence from behavioral and electrophysiologic measures. *J. Am. Acad. Audiol.* 4, 122–131.

Karaer, I., and Gorkem, T. (2020). The effect of premature ovarian failure on inner ear function. *J. Obstet. Gynaecol.* 40, 247–251. doi: 10.1080/01443615.2019.1621815

Kemp, D. (1978). Stimulated acoustic emissions from within the human auditory system. *J. Acoust. Soc. Am.* 64, 1386–1391. doi: 10.1121/1.382104

Kemp, D. (2002). Otoacoustic emissions, their origin in cochlear function, and use. *Br. Med. Bull.* 63, 223–241. doi: 10.1093/bmb/63.1.223

Kenny, R., Patil, N., and Considine, N. (2011). Sudden (reversible) sensorineural hearing loss in pregnancy. *Ir. J. Med. Sci.* 180, 79–84. doi: 10.1007/s11845-010-0525-z

Kilicdag, E., Yavuz, H., Bagis, T., Tarim, E., Erkan, A., and Kazanci, F. (2004). Effects of estrogen therapy on hearing in postmenopausal women. *Am. J. Obstet. Gynecol.* 190, 77–82. doi: 10.1016/j.ajog.2003.06.001

Kim, S. H., Kang, B. M., Chae, H. D., and Kim, C. H. (2002). The association between serum estradiol level and hearing sensitivity in postmenopausal women. *Obstet. Gynecol.* 99(5 Pt 1), 726–730. doi: 10.1016/s0029-7844(02)01963-4

Kim, S., Lim, E., Kim, H., Park, J., Jarng, S., and Lee, S. (2010). Sex differences in a cross sectional study of age-related hearing loss in Korean. *Clin Exp Otorhinolaryngol.* 3, 27–31. doi: 10.3342/ceo.2010.3.1.27

Köşüş, N., Köşüş, A., Turhan, N., and Kurtaran, H. (2012). Hearing levels in menopausal women and the effect of tibolone on audiological functions. *J. Obstet. Gynaecol.* 32, 294–297. doi: 10.3109/01443615.2011.649316

Kuijper, E., Ket, J., Caanen, M., and Lambalk, C. (2013). Reproductive hormone concentrations in pregnancy and neonates: a systematic review. *Reprod. Biomed. Online* 27, 33–63. doi: 10.1016/j.rbmo.2013.03.009

Laugel, G., Dengerink, H., and Wright, J. (1987). Ovarian steroid and vasoconstrictor effects on cochlear blood flow. *Hear. Res.* 31, 245–251. doi: 10.1016/0378-5955(87)90194-8

Lauretta, R., Sansone, M., Sansone, A., Romanelli, F., and Appetecchia, M. (2018). Gender in endocrine diseases: role of sex gonadal hormones. *Int. J. Endocrinol.* 2018:4847376. doi: 10.1155/2018/4847376

Lee, J., and Marcus, D. (2001). Estrogen acutely inhibits ion transport by isolated stria vascularis. *Hear. Res.* 158, 123–130. doi: 10.1016/s0378-5955(01)00316-1

Lee, S.-S., Han, K.-D., and Joo, Y.-H. (2017). Association of perceived tinnitus with duration of hormone replacement therapy in Korean postmenopausal women: a cross-sectional study. *BMJ Open* 7, e013736. doi: 10.1136/bmjopen-2016-013736

Liu, J., Wang, D., Li, X., and Ningyu, W. (2017). Association between sex and speech auditory brainstem responses in adults, and relationship to sex hormone levels. *Med. Sci. Monit.* 23, 2275–2283. doi: 10.12659/msm.904651

Mann, N., Sidhu, R., and Babbar, R. (2012). Brainstem auditory evoked responses in different phases of menstrual cycle. *J. Clin. Diagn. Res.* 6, 1640–1643. doi: 10.7860/JCDR/2012/4896.2621

Manyonda, I., Sinai Talaulikar, V., Pirhadi, R., Ward, J., Banerjee, D., and Onwude, J. (2022). Could perimenopausal estrogen prevent breast cancer? Exploring the differential effects of estrogen-only versus combined hormone replacement therapy. *J. Clin. Med. Res.* 14, 1–7. doi: 10.14740/jocmr4646

McFadden, D. (1993). A speculation about the parallel ear asymmetries and sex differences in hearing sensitivity and otoacoustic emissions. *Hear. Res.* 68, 143–151. doi: 10.1016/0378-5955(93)90118-k

McFadden, D. (1998). Sex differences in the auditory system. *Dev Neuropsychol.* 14, 261–298.

McFadden, D., Pasanen, E., Raper, J., Wallen, K., Weldele, M., and Glickman, S. (2005). Otoacoustic emissions measured in rhesus monkeys (*Macaca mulatta*) and spotted hyenas (*Crocuta Crocuta*): effects of sex, season, and prenatal hormone exposure. *Horm. Behav.* 48, 114–114.

McFadden, D., Pasanen, E., Raper, J., Lange, H., and Wallen, K. (2006). Sex differences in otoacoustic emissions measured in rhesus monkeys (*Macaca mulatta*). *Horm. Behav.* 50, 274–284. doi: 10.1016/j.yhbeh.2006.03.012

Meltser, I., Tahera, Y., Simpson, E., Hultcrantz, M., Charitidi, K., Gustafsson, J., et al. (2008). Estrogen receptor beta protects against acoustic trauma in mice. *J. Clin. Invest.* 118, 1563–1570. doi: 10.1172/JCI32796

Melynyte, S., Pipinis, E., Genyte, V., Voicikas, A., Rihs, T., and Griskova-Bulanova, I. (2018). 40 Hz auditory steady-state response: the impact of handedness and gender. *Brain Topogr.* 31, 419–429. doi: 10.1007/s10548-017-0611-x

Milon, B., Mitra, S., Song, Y., Margulies, Z., Casserly, R., Drake, V., et al. (2018). The impact of biological sex on the response to noise and otoprotective therapies against acoustic injury in mice. *Biol. Sex. Differ.* 9:12. doi: 10.1186/s13293-018-0171-0

Mitre, E., Figueira, A., Rocha, A., and Alves, S. (2006). Audiometric and vestibular evaluation in women using the hormonal contraceptive method. *Braz. J. Otorhinolaryngol.* 72, 350–354. doi: 10.1016/s1808-8694(15)30967-8

Moenter, S., Chu, Z., and Christian, C. (2009). Neurobiological mechanisms underlying oestradiol negative and positive feedback regulation of gonadotrophin-releasing hormone neurones. *J. Neuroendocrinol.* 21, 327–333. doi: 10.1111/j.1365-2826.2009.01826.x

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., et al. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst. Rev.* 4:1. doi: 10.1186/2046-4053-4-1

Morimoto, N., Tanaka, T., Taiji, H., Horikawa, R., Naiki, Y., Morimoto, Y., et al. (2006). Hearing loss in Turner syndrome. *J. Pediatr.* 149, 697–701. doi: 10.1016/j.jpeds.2006.06.071

Mościcki, E. K., Elkins, E. F., Baum, H. M., and McNamara, P. M. (1985). Hearing loss in the elderly: an epidemiologic study of the Framingham heart study cohort. *Ear. Hear.* 6, 184–190.

Motohashi, R., Takumida, M., Shimizu, A., Konomi, U., Fujita, K., Hirakawa, K., et al. (2010). Effects of age and sex on the expression of estrogen receptor alpha and beta in the mouse inner ear. *Acta Otolaryngol.* 130, 204–214. doi: 10.3109/00016480903016570

Najmabadi, S., Schliep, K., Simonsen, S., Porucznik, C., Egger, M., and Stanford, J. (2020). Menstrual bleeding, cycle length, and follicular and luteal phase lengths in women without known subfertility: a pooled analysis of three cohorts. *Paediatr. Perinat. Epidemiol.* 34, 318–327. doi: 10.1111/ppe.12644

Newmark, M., Merlob, P., Bresloff, I., Olsha, M., and Attias, J. (1997). Click evoked otoacoustic emissions: inter-aural and gender differences in newborns. *J. Basic Clin. Physiol. Pharmacol.* 8, 133–139. doi: 10.1515/jbcpp.1997.8.3.133

NIHR, (2020). *Prospero international prospective register of systematic reviews*. Available online at: https://www.crd.york.ac.uk/PROSPERO/ (accessed November 11, 2020).

O'Connell, L., Ding, J., Ryan, M., and Hofmann, H. (2011). Neural distribution of the nuclear progesterone receptor in the Túngara frog, Physalaemus pustulosus. *J. Chem. Neuroanat.* 41, 137–147. doi: 10.1016/j.jchemneu.2011.01.002

Page, M., McKenzie, J., Bossuyt, P., Boutron, I., Hoffmann, T., Mulrow, C., et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 372:n71. doi: 10.1136/bmj.n71

Pearson, J., Morrell, C., Gordon-Salant, S., Brant, L., Metter, E., Klein, L., et al. (1995). Gender differences in a longitudinal study of age-associated hearing loss. *J Acoust. Soc. Am.* 97, 1196–1205. doi: 10.1121/1.412231

Penner, M., and Zhang, T. (1997). Prevalence of spontaneous otoacoustic emissions in adults revisited. *Hear. Res.* 103, 28–34. doi: 10.1016/s0378-5955(96)00162-1

Price, K., Zhu, X., Guimaraes, P., Vasilyeva, O., and Frisina, R. (2009). Hormone replacement therapy diminishes hearing in peri-menopausal mice. *Hear. Res.* 252, 29–36. doi: 10.1016/j.heares.2009.02.010

Robinette, M., and Glattke, T. (1997). *Otoacoustic emissions: clinical applications*, 3rd Edn. New York, NY: Thieme.

Rosenhamer, H., Lindström, B., and Lundborg, T. (1980). On the use of click-evoked electric brainstem responses in audiological diagnosis. II. The influence of sex and age upon the normal response. *Scand. Audiol.* 9, 93–100. doi: 10.3109/01050398009076342

Sennaroglu, G., and Belgin, E. (2001). Audiological findings in pregnancy. *J. Laryngol. Otol.* 115, 617–621. doi: 10.1258/0022215011908603

Serra, A., Maiolino, L., Agnello, C., Messina, A., and Caruso, S. (2003). Auditory brain stem response throughout the menstrual cycle. *Ann. Otol. Rhinol. Laryngol.* 112, 549–553. doi: 10.1177/000348940311200612

Sharashenidze, N., Schacht, J., and Kevanishvili, Z. (2008). Age-related hearing loss: Gender differences. *Georgian Med. News* 144, 14–18.

Shuster, B., Depireux, D., Mong, J., and Hertzano, R. (2019). Sex differences in hearing: probing the role of estrogen signaling. *J. Acoust. Soc. Am.* 145:3656. doi: 10.1121/1.5111870

Simonoska, R., Stenberg, A., Masironi, B., Sahlin, L., and Hultcrantz, M. (2009). Estrogen receptors in the inner ear during different stages of pregnancy and development in the rat. *Acta Otolaryngol.* 129, 1175–1181. doi: 10.3109/00016480802691150

Sisneros, J., Bass, A., and Sisneros, J. (2003). Seasonal plasticity of peripheral auditory frequency sensitivity. *J. Neurosci.* 23, 1049–1058.

Sisneros, J., Forlano, P., Deitcher, D., and Bass, A. (2004). Steroid-dependent auditory plasticity leads to adaptive coupling of sender and receiver. *Science* 305, 404–407. doi: 10.1126/science.1097218

Snihur, A. W. K., and Hampson, E. (2011). Sex and ear differences in spontaneous and click-evoked otoacoustic emissions in young adults. *Brain Cogn.* 77, 40–47.

Souza, D., Luckwu, B., Andrade, W., Pessoa, L., Nascimento, J., and Rosa, M. (2017). Variation in the hearing threshold in women during the menstrual cycle. *Int Arch. Otorhinolaryngol.* 21, 323–328. doi: 10.1055/s-0037-1598601

Stenberg, A., Nylén, O., Windh, M., and Hultcrantz, M. (1998). Otological problems in children with Turner's syndrome. *Hear. Res.* 124, 85–90. doi: 10.1016/s0378-5955(98)00113-0

Stenberg, A., Simonoska, R., Stygar, D., Sahlin, L., and Hultcrantz, M. (2003). Effect of estrogen and antiestrogens on the estrogen receptor content in the cochlea of ovariectomized rats. *Hear. Res.* 182, 19–23. doi: 10.1016/s0378-5955(03)00136-9

Stenberg, A., Wang, H., Fish, J., Schrott-Fischer, A., Sahlin, L., and Hultcrantz, M. (2001). Estrogen receptors in the normal adult and developing human inner ear and in Turner's syndrome. *Hear. Res.* 157, 87–92. doi: 10.1016/s0378-5955(01)00280-5

Stenberg, A., Wang, H., Sahlin, L., and Hultcrantz, M. (1999). Mapping of estrogen receptors α and β in the inner ear of mouse and rat. *Hear. Res.* 136, 29–34.

Stuart, A., and Kerls, A. (2018). Does contralateral inhibition of transient evoked otoacoustic emissions suggest sex or ear laterality effects? *Am. J. Audiol.* 27, 272–282. doi: 10.1044/2018_AJA-17-0106

Svechnikov, K., and Söder, O. (2007). Ontogeny of gonadal sex steroids. *Best Pract. Res. Clin. Endocrinol. Metab.* 22, 95–106. doi: 10.1016/j.beem.2007.09.002

Svedbrant, J., Bark, R., Hultcrantz, M., and Hederstierna, C. (2015). Hearing decline in menopausal women–a 10-year follow-up. *Acta Otolaryngol.* 135, 807–813. doi: 10.3109/00016489.2015.1023354

Talmadge, C., Long, G., Murphy, W., and Tubis, A. (1993). New off-line method for detecting spontaneous otoacoustic emissions in human subjects. *Hear. Res.* 71, 170–182. doi: 10.1016/0378-5955(93)90032-v

Tandon, O., Khaliq, F., and Goel, N. (2001). Auditory evoked potential responses in menopausal women: a normative study. *Indian J. Physiol. Pharmacol.* 45, 361–366.

Tremere, L., Burrows, K., Jeong, J., and Pinaud, R. (2011). Organization of estrogen-associated circuits in the mouse primary auditory cortex. *J. Exp. Neurosci.* 2011, 45–60. doi: 10.4137/JEN.S7744

Tremere, L., Jeong, J., Pinaud, R., and Tremere, L. (2009). Estradiol shapes auditory processing in the adult brain by regulating inhibitory transmission and plasticity-associated gene expression. *J. Neurosci.* 29, 5949–5963. doi: 10.1523/JNEUROSCI.0774-09.2009

Trott, S., Cline, T., Weihing, J., Beshear, D., Bush, M., and Shinn, J. (2019). Hormones and hearing: central auditory processing in women. *J. Am. Acad. Audiol.* 30, 493–501. doi: 10.3766/jaaa.17123

Tyagi, V., Scordo, M., Yoon, R., Liporace, F., and Greene, L. (2017). Revisiting the role of testosterone: are we missing something? *Rev Urol.* 19, 16–24. doi: 10.3909/riu0716

Upadhayay, N., Paudel, B., Singh, P., Bhattarai, B., and Agrawal, K. (2014). Pre- and postovulatory auditory brainstem response in normal women. *Indian J. Otolaryngol. Head Neck Surg.* 66(Suppl. 1), 133–137. doi: 10.1007/s12070-011-0378-4

Villavisanis, D., Schrode, K., and Lauer, A. (2018). Sex bias in basic and preclinical age-related hearing loss research. *Biol. Sex Diff.* 9:23. doi: 10.1186/s13293-018-0185-7

Wadnerkar, M., Whiteside, S., and Cowell, P. (2008). Dichotic listening asymmetry: sex differences and menstrual cycle effects. *Lateral. Asymmetr. Body Brain Cogn.* 13, 297–309.

Walpurger, V., Pietrowsky, R., Kirschbaum, C., and Wolf, O. (2004). Effects of the menstrual cycle on auditory event-related potentials. *Horm. Behav.* 46, 600–606. doi: 10.1016/j.yhbeh.2004.07.002

Wang, Q., Wang, X., Yang, L., Han, K., Huang, Z., and Wu, H. (2021). Sex differences in noise-induced hearing loss: a cross-sectional study in China. *Biol. Sex Differ.* 12:24. doi: 10.1186/s13293-021-00369-0

Wells, G., Shea, B., O'connell, D., Peterson, J., Welch, V., Losos, M., et al. (2000). *The Newcastle-Ottawa scale (Nos) for assessing the quality of nonrandomised studies in meta-analyses*. Ottawa, ON: Ottawa Hospital Research Institute.

Wharton, J., and Church, G. (1991). Influence of menopause on the auditory brainstem response. *Maturitas* 13, 164–164.

Yadav, A., Tandon, O. P., and Vaney, N. (2003). Long latency auditory evoked responses in ovulatory and anovulatory menstrual cycle. *Indian J. Physiol. Pharmacol.* 47, 179–184.

Zakaria, M. N., Abdul Wahab, N. A., Maamor, N., Jalaei, B., and Dzulkarnain, A. A. A. (2019). Auditory brainstem response (ABR) findings in males and females with comparable head sizes at supra-threshold and threshold levels. *Neurol. Psychiatry Brain Res.* 32, 4–7.

Zhang, J., Zhang, T., Yu, L., Ruan, Q., Yin, L., Liu, D., et al. (2018). Effects of ovarian reserve and hormone therapy on hearing in premenopausal and postmenopausal women: a cross-sectional study. *Maturitas* 111, 77–81. doi: 10.1016/j.maturitas.2018.01.019

Zündorf, I. C., Karnath, H.-O., and Lewald, J. (2011). Male advantage in sound localization at cocktail parties. *Cortex* 47, 741–749.

frontiers | Frontiers in Human Neuroscience

TYPE Review
PUBLISHED 17 May 2023
DOI 10.3389/fnhum.2023.1181760

OPEN ACCESS

EDITED BY
Alessia Tonelli,
Italian Institute of Technology (IIT), Italy

REVIEWED BY
Ambra Ferrari,
Max Planck Institute for Psycholinguistics,
Netherlands
Cristiano Cuppini,
University of Bologna, Italy
Patrick Bruns,
University of Hamburg, Germany

*CORRESPONDENCE
Ladan Shams
✉ lshams@psych.ucla.edu

RECEIVED 07 March 2023
ACCEPTED 02 May 2023
PUBLISHED 17 May 2023

CITATION
Murray CA and Shams L (2023) Crossmodal interactions in human learning and memory. *Front. Hum. Neurosci.* 17:1181760.
doi: 10.3389/fnhum.2023.1181760

Crossmodal interactions in human learning and memory

Carolyn A. Murray[1] and Ladan Shams[1,2]*

[1]Department of Psychology, University of California, Los Angeles, Los Angeles, CA, United States,
[2]Department of Bioengineering, Neuroscience Interdepartmental Program, University of California, Los Angeles, Los Angeles, CA, United States

Most studies of memory and perceptual learning in humans have employed unisensory settings to simplify the study paradigm. However, in daily life we are often surrounded by complex and cluttered scenes made up of many objects and sources of sensory stimulation. Our experiences are, therefore, highly multisensory both when passively observing the world and when acting and navigating. We argue that human learning and memory systems are evolved to operate under these multisensory and dynamic conditions. The nervous system exploits the rich array of sensory inputs in this process, is sensitive to the relationship between the sensory inputs, and continuously updates sensory representations, and encodes memory traces based on the relationship between the senses. We review some recent findings that demonstrate a range of human learning and memory phenomena in which the interactions between visual and auditory modalities play an important role, and suggest possible neural mechanisms that can underlie some surprising recent findings. We outline open questions as well as directions of future research to unravel human perceptual learning and memory.

KEYWORDS

multisensory, perceptual learning, adaptation, recalibration, multisensory memory, multisensory learning

1. Introduction

The environment and set of tasks the human brain must complete throughout the course of our lives create an immense challenge for the nervous system. We live in dynamic environments, whose changes require a large variety of flexible behaviors to navigate. Moreover, the human body also changes through time, growing when we are young and deteriorating with age. The brain must recalibrate and adjust its functioning during all of these stages in life. The complexity of these systems is such that it is not possible for all behaviors to be hard-coded; the human genome only contains 20–25 thousand genes, which is far too few to code everything the brain must compute and perform. In addition, humans are social animals, which will require us to not just have a functional understanding of our physical environment, but of our social experiences and networks as well.

These complex environmental and developmental factors have thus necessitated the evolution of a brain that is capable of recalibration and learning. The human brain is, in fact, noted for being incredibly plastic (Kolb and Whishaw, 1998; Calford, 2002), and apt at both supervised and unsupervised learning (Knudsen, 1994). In addition, the human brain is accomplished in memory tasks that support learning about our environments and remembering our social interactions. As they are such fundamental functions of human behavior, both learning and memory have been studied extensively in humans over the decades in a variety of disciplines and using a variety of methods. However, the vast majority of these studies focus on studying one sense at a time [for overviews, see Goldstone (1998), Fiser and Lengyel (2022)].

While situations that focus on the experiences of only one sense can be created in an experimental space, such work does not reflect the cues across many senses that would be available and working in concert in a natural environment. On a daily basis, we use information across multiple senses to learn about our environment and encode in our memories for later use. The senses do not operate in a vacuum. If we drop a glass, we do not just see it fall, but we hear the impact and feel the lack of its weight in our hands. When talking to a friend, we do not just hear their voice, but see their facial expressions and smell their perfume. With such rich information available across senses about the same experience, it would make sense if the brain was capable of processing this information in a holistic way, without the boundaries of sensory modality and perhaps even exploiting the relationship between the sensory cues. Yet, the vast majority of studies of perceptual learning and memory have used unisensory stimuli and tasks.

Research over the last two decades, however, has greatly enhanced our understanding of how the brain is able to combine information across the senses. Myriad studies have established that sensory pathways can influence one another, even at their earliest stages. For example, the presence of low-level multisensory illusions, such as the ventriloquist illusion (Thurlow and Jack, 1973; Bruns, 2019) and the sound-induced flash illusion (Shams et al., 2000; Hirst et al., 2020) indicate that the senses combine information early on and influence one another in ways that are observable at a behavioral level. Psychophysical studies have established that the interactions between the senses is ubiquitous, they occur across all sensory modalities and many tasks (e.g., Botvinick and Cohen, 1998; Shams et al., 2000; Wozny et al., 2008; Peters et al., 2015; Bruns, 2019), and across the lifespan (e.g., Setti et al., 2011; Burr and Gori, 2012; Nardini and Cowie, 2012; Murray et al., 2016a; McGovern et al., 2022). Accordingly, brain studies have revealed interactions between the senses at a variety of processing stages, in all processing domains (Murray et al., 2016b; Ferraro et al., 2020; Gau et al., 2020, and see Ghazanfar and Schroeder, 2006; Driver and Noesselt, 2008; for reviews). Altogether, research has uncovered that multisensory processing is not simply the sum of unisensory processes, which implies that multisensory learning cannot be simplified to the sum of the constituent unisensory learning and memory. Indeed, researchers have begun investigating learning and memory under multisensory conditions, and these studies have revealed surprising phenomena that point to multisensory processing being a unique and powerful mechanism for learning and memory.

Here, we will briefly review some of the studies that investigate learning and memory through a multisensory lens, with a particular focus on audio-visual studies. We will additionally focus on studies performed in healthy human adults, though there is significant work studying multisensory learning during development (e.g. Gori et al., 2008; Nardini and Cowie, 2012; Dionne-Dostie et al., 2015; Murray et al., 2016a), in clinical populations (e.g., Held et al., 2011; Landry et al., 2013; Stevenson et al., 2017), and in animals (e.g., Wallace et al., 2004; Xu et al., 2014). We will highlight key takeaways from healthy human adult research as a whole. Building upon neural mechanisms proposed by Shams and Seitz (2008), we will outline possible neural mechanisms that may explain the relative potency of multisensory learning/memory when compared to unisensory variations, and a larger range of learning phenomena including some surprising recent behavioral findings. We additionally suggest directions for future research.

2. Multisensory learning

The topic of multisensory learning has been broadly approached under a number of labels, including but not limited to studies of multimedia learning (Mayer, 2014) or Montessori education (Montessori, 2013). However, many of these studies, by nature of being more applied in nature, are often not rigorous experiments with appropriate controls. Thus, the results are frequently not easy to interpret. In our discussion of multisensory learning, we will focus on experimental studies that, in addition to using rigorous experimental methods, also shed light on underlying mechanisms that could explain multisensory benefits. These studies have tackled a variety of learning ranging from supervised perceptual learning to unsupervised or implicit types of learning such as recalibration and adaptation.

2.1. Perceptual learning

Perceptual learning can be defined as a refinement in perceptual processes, improving detection and discrimination of stimuli through perceptual experience (Gold and Watanabe, 2010). Because the experience is crucial for improvement, there has been significant interest in developing training regimens that will support perceptual learning. Sensory training has been long studied in unisensory contexts (for examples, see reviews by Goldstone, 1998; Fiser and Lengyel, 2022). However, studies in multisensory perceptual learning have emerged in the past two decades that indicate this learning is not solely a unisensory phenomenon, and that multisensory training has the potential to be a powerful tool for refining perception above and beyond that obtained by unisensory training.

One fascinating benefit of multisensory training is the ability for this sensory information to refine not just multisensory processing, but to improve on unisensory processing. In the domain of motion processing, audio-visual training has been shown to be superior to visual training both in the overall degree of learning as well as rate of learning, even when compared on trials consisting only of visual information (Seitz et al., 2006). Furthermore, a later study (Kim et al., 2008) showed that the congruence between the auditory and the visual motion during training was necessary for this multisensory training benefit. Training with incongruent audiovisual stimuli did not lead to improved learning compared to visual-alone training, even though the stimuli in the incongruent condition were equally arousing as those in the congruent condition. These results suggest that integration of auditory-visual stimuli is critical for the facilitation and enhancement of learning, making the benefit a matter of multisensory mechanisms being used, rather than a mere effect of heightened neural activity due to potentially increased arousal.

In this study, the participants in the multisensory training groups were trained with sessions that consisted of mostly auditory-visual trials, however, it also included some visual-only trials. This design also allowed comparing the accuracy in unisensory versus multisensory trials for each subject throughout training. **Figure 1** shows the detection accuracy for the congruent auditory-visual training group for both auditory-visual trials (broken green line) and visual-alone trials (solid green line). In auditory-visual trials, there is task-relevant information (i.e., which of the two intervals contains coherent motion) in both modalities, whereas in the visual-alone trials that information is only available in the visual modality. The coherence level of visual stimuli were equivalent between visual-alone and auditory-visual trials. Therefore, it was expected that performance in auditory-visual trials to be higher than that of visual-only trials. Indeed, in the early training sessions, participants' performance was higher in the audiovisual trials than in visual trials. However, this difference decreased over subsequent training sessions, and finally the performance in the visual-only trials matched that of auditory-visual trials by the end of training (**Figure 1**). This intriguing finding has important implications for unraveling the computational mechanisms of multisensory learning as we discuss later.

Work by von Kriegstein and Giraud (2006) showed that neural changes that occurred during multisensory learning could explain such phenomena. Training individuals on audiovisual voice-face associations strengthened the functional connection between face- and voice-recognition regions of the brain. They argue that this means that multisensory training has the means to improve unisensory perceptual improvement because later unisensory representations have the ability to activate larger ensembles due to increased connectivity through multisensory training. To that end, multisensory training has the ability to be more effective for perceptual learning than unisensory alternatives, perhaps as a result of multisensory mechanisms that will be discussed in more depth in the Neural Mechanisms section below.

In a more recent study, Barakat et al. (2015), investigated the multisensory training benefit in the context of rhythm perception. Participants were asked to make same/different judgments on visual rhythms. Participants were trained in either a visual only condition, an auditory only condition, or a multisensory condition, where identical auditory and visual rhythms were played simultaneously. In line with previous findings, but even more strikingly, they found that participants who underwent the multisensory training improved in the visual task substantially and already after one training session, in contrast to the participants who underwent visual-only training who showed no significant improvement even after two training sessions. Perhaps more surprising, however, was the finding that the auditory training was as effective as multisensory training, even though sound was completely absent in the test task. This pattern of results suggests that the visual and auditory regions must be communicating with one another even in the absence of a multisensory training, meaning crossmodal mechanisms must be engaged even in the absence of direct stimulation.

These findings are consistent with those of a more recent study that examined crossmodal transfer of learning in both spatial and temporal tasks in both vision and hearing (McGovern et al., 2016). The results showed that in a given task training in sensory modality that is relatively more accurate (e.g., vision in a spatial task, hearing in a temporal task) leads to improved performance of the less accurate sensory modality in the same task. Such findings cannot be easily explained by traditional theories of perceptual learning. Possible neural mechanisms mediating these phenomena will be further explored later in the (section "4. Neural mechanisms").

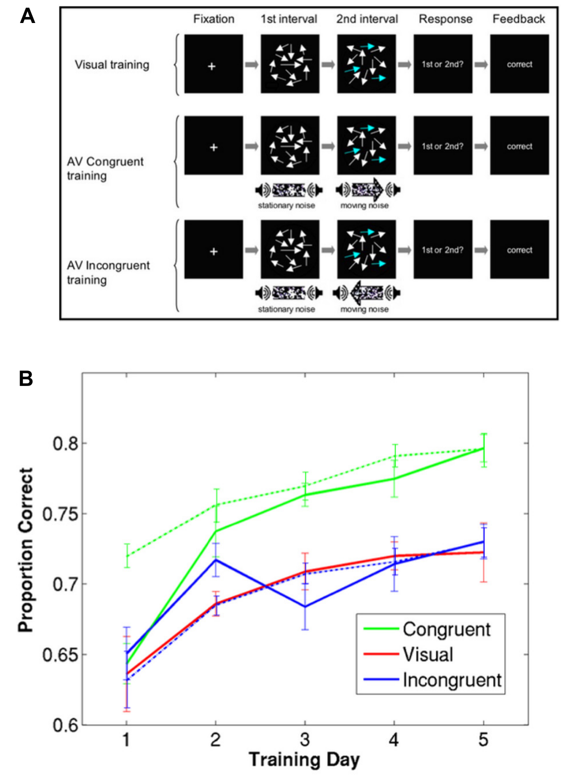


FIGURE 1
Perceptual learning of motion coherence detection in three different training conditions, adapted from Kim et al. (2008). **(A)** Multisensory training and visual training and test conditions presented to participants in Kim et al. (2008). **(B)** A group of participants was trained using only visual stimuli (red), a group was trained using visual and auditory stimuli moving in the same direction (green), and a group was trained using visual and auditory stimuli moving in opposite directions (blue). A fraction of trials in the multisensory training groups were unisensory (visual only). The solid lines and broken lines represent performance on visual-only trials and audiovisual trials, respectively. Error bars represent within-group standard errors ($n = 7$).

While the aforementioned studies have trained observers on performing a perceptual task that can be done using both unisensory and multisensory stimuli (e.g., detecting motion), other studies have investigated the effect of training observers on a task involving determination of the temporal relationship between crossmodal stimuli, namely, the simultaneity or the temporal order of two crossmodal stimuli (e.g., Powers et al., 2009; Alais and Cass, 2010; see O'Brien et al., 2023 for a recent review). These studies have reported improved performance on the trained tasks (e.g., Virsu et al., 2008; Powers et al., 2009; Alais and Cass, 2010; De Niear et al., 2018), and in some cases also a transfer of learning to other tasks involving crossmodal stimuli (e.g., Setti et al., 2014; McGovern et al., 2016; Powers et al., 2016; Sürig et al., 2018, but see Horsfall et al., 2021; O'Brien et al., 2020). These findings demonstrate the fast plasticity of the perceptual processes even at foundational level of time representation. However, the exact mechanism underlying the improved performance (i.e., narrowing of the time window of simultaneity or improved temporal acuity) requires further research. Improved performance in these tasks could be due to either the improved unisensory temporal precision, or a modification of multisensory mechanisms, or both. Future research can elucidate this by testing observers in unisensory tasks before and after training, and/or using the Bayesian Causal Inference model to quantitatively probe the unisensory precisions as well as multisensory processing components before and after training.

2.2. Recalibration

While perceptual learning studies typically involve giving feedback to the participants about the accuracy of their responses, and therefore are a form of supervised learning, other types of learning that occur naturally in nature and do not involve explicit feedback also play an important role in being able to function in an ever-changing environment. For example, the brain needs to be able to maintain coherence of information across the senses. Were the senses truly independent, it wouldn't be possible to use one to calibrate another. Thus, crossmodal interactions are also critical in maintaining the accuracy of sensory measurements and representations in face of environmental and bodily changes. It is well established that the human nervous system is capable of recalibrating the sensory systems even in maturity in various processing domains (e.g., Recanzone, 1998; Fujisaki et al., 2004; Vroomen et al., 2004). For example, repeated exposure to auditory-visual stimuli with a fixed spatial discrepancy leads to a subsequent shift in the map of auditory space in the direction of the previously experienced visual stimuli, in a phenomenon known as the *ventriloquist aftereffect* (Recanzone, 1998). This is a clear illustration of the use of the visual input as a teaching signal to calibrate the auditory representations. Indeed, quantitatively modeling the observer's localization responses before and after exposure to spatially discrepant auditory-visual stimuli has shown that it is the sensory (namely, auditory) representations that are shifted in ventriloquist aftereffect rather than a prior expectation of stimuli or a combination of the two (Wozny and Shams, 2011).

While earlier studies had utilized extended exposure (hundreds or thousands of trials, or minutes or hours of exposure), a more recent study (Wozny and Shams, 2011) showed that long exposure is not required to trigger and engage the recalibration process. A single exposure lasting only a fraction of a second to a spatially discrepant audiovisual stimulus can cause a shift in spatial localization of an ensuing auditory stimulus presented alone (Wozny and Shams, 2011). Recalibration in the span of a fraction of a second indicates that the nervous system is extremely sensitive to discrepancy across senses and seeks to resolve it expeditiously. Because multisensory stimuli can be used in such rapid recalibration, they are uniquely poised as crucial to help the brain to keep up with a dynamic environment. The effects of recalibration can be long-lasting, to match the environment; for example, multisensory recalibration in the ventriloquist aftereffect has been shown to persist over the course of days, with appropriate training (Bruns, 2019).

While recalibration has been studied extensively both at a behavioral and neural level in both humans and animal models (for example, Knudsen and Knudsen, 1985; Wallace et al., 1998; Kopco et al., 2009; Aller et al., 2022) the computational characterization of this process had not been investigated systematically until recently. Wozny and Shams (2011) probed the role of causal inference in the visual recalibration of auditory space in the same study. Recalibration seemed significantly stronger on trials where observers appeared to have inferred a common cause for the auditory and visual stimuli compared to those where did not appear to perceive unity. Auditory recalibration by vision also appears to be better explained by Bayesian Causal Inference than by competing models of sensory reliability or fixed-ratio recalibration (Hong et al., 2021). Such findings are surprising because recalibration is traditionally considered a very low-level phenomenon, occurring at early stages of sensory processing [as in Zwiers et al. (2003); Fujisaki et al. (2004)], whereas causal inference is considered a high-level process, occurring in later stages of cortical processing (Kayser and Shams, 2015; Rohe and Noppeney, 2015; Aller and Noppeney, 2019; Cao et al., 2019; Rohe et al., 2019; Ferrari and Noppeney, 2021). Recent works are challenging this distinction, however; it has been recently suggested that recalibration can be subject to top-down influences from higher cognitive processes (Kramer et al., 2020), and that regions involved in both perception and decision-making are flexibly involved in the recalibration process (Aller et al., 2022). Such findings support the computational evidence that low-level perceptual and higher-level computational processes may not be as distinct as originally theorized, and therefore, causal inference could influence the recalibration process.

2.3. Implicit associative learning

Implicit associative learning is another form of unsupervised learning, where a new association is learned based on passive exposure to statistical regularities of the environment (Reber, 1967; Knowlton et al., 1994; Saffran et al., 1996; Aslin, 2017; Batterink et al., 2019; Sherman et al., 2020). Observers are able to implicitly learn the association between crossmodal stimuli, even when the association is entirely arbitrary. For example, exposure to arbitrary association between visual brightness and haptic stiffness results in refined discrimination of visual brightness (Ernst, 2007).

Because this type of learning involves extraction of statistical regularities in the environment it falls under the umbrella of statistical learning, broadly speaking. Statistical learning has been studied often from a unisensory perspective (Conway and Christiansen, 2005), but studies that have examined statistical learning across sensory modalities have often reported a powerful and fast learning of links (joint or conditional probabilities) between the senses, such as shape and sound (Seitz et al., 2007). In a study that compared the rate of learning of within-modality regularities vs. across-modality regularities, it was found that observers learned auditory-visual regularities more effectively than visual-visual or auditory-auditory ones (Seitz et al., 2007). Therefore, it appears that the nervous system is particularly apt at detecting statistical relationships across the senses. However, there may be constraints on temporal relationships that lend themselves to learning of crossmodal statistical regularities. Many studies showing multisensory benefit in implicit association tasks utilize simultaneous audiovisual presentation, but some studies indicate that learning multisensory associations through time, including between color and tone (Conway and Christiansen, 2006) or crossmodal artificial grammar sequences (Walk and Conway, 2016) may be more challenging to learn than within-modality associations. Such findings potentially suggest there may be limitations to the types of procedures that will produce effective multisensory learning. Such suggestions do not preclude that multisensory learning is possible, just that the constraints on this learning may be different from those on unisensory learning (Frost et al., 2015). The necessity of crossmodal synchronicity for effective implicit associative multimodal learning is thus an open question in need of more research.

It should also be noted that, as with other forms of learning discussed earlier, benefits can be observed even when one modality present during learning is irrelevant at test. In a study in which participants were passively exposed to co-occurring visual and auditory features in the background, and in a subsequent visual test, they exhibited improved sensitivity to visual features in presence of the associated sound, even though the sound was task-irrelevant (Shams et al., 2011). Altogether, these findings highlight that multisensory encoding of information is able to improve unisensory representation and processing, even if the relationship between the two stimuli in different senses is arbitrary.

In fact, learning associations that are seemingly arbitrary could be a crucial step in learning meaningful associations. Learning of crossmodal correspondences– information across senses that are arbitrary yet are robustly considered “congruent”– are an important area of study within multisensory processing (for reviews, see Spence, 2011; Parise, 2016). Such correspondences have been studied across a wide variety of sensory pairs, including auditory timbre and visual properties such as shape and color (Adeli et al., 2014), haptic assessment of heaviness and auditory pitch (Walker et al., 2017), visual hue and tactile texture (Jraissati et al., 2016), and visual color and gustatory taste profile of an object (Spence et al., 2010). While these associations range from the seemingly sensible to the entirely arbitrary, they usually evolve from some type of association present in the environment to some extent (for discussion, see Parise, 2016), and thus reflect a great flexibility in crossmodal learning in order to map such seemingly arbitrary associations. While the crossmodal correspondence is rightly treated as related yet separate from a truly multisensory process, current research indicates that crossmodal correspondences, once learned, can influence multisensory integration. Training in an arbitrary but “congruent” crossmodal correspondences has been shown to prime later multisensory integration (Brunel et al., 2015), and as such may represent a crucial stage in understanding how the brain learns to integrate novel crossmodal pairs. The neural mechanisms by which such crossmodal correspondences develop and persist remain unclear; though it has been posited that they may be the same mechanisms that underlie the phenomenon of synesthesia (Parise and Spence, 2009), further research into the mechanisms investigating how crossmodal correspondences contribute to multisensory integration are required.

3. Multisensory memory

The benefits of multisensory processing are not limited to just the realm of learning. The memory systems of the brain must also, crucially, be able to store and represent information across senses in order for humans to make sense of our environment. In addition, our episodic memory, as well as being a useful guide on our environment, helps us to store information crucial to the events of our lives, which helps us to store information crucial to social interactions and aid in decision making critical for survival. Episodic memory is commonly defined as memories for events and experiences, rich in sensory and contextual details, rather than memories for facts (Tulving, 1993). Memories are rich in sensory detail and can typically be cured by many senses. Neuroimaging studies have revealed that the role of perception in memory was not unidirectional upon encoding: recall of visual and auditory stimuli reactivates sensory-specific cortices that were active at encoding. This is true within modality, where a sensory region active during encoding is reactivated upon recall (Nyberg et al., 2000) but has also been shown in multisensory conditions, where a visual probe for an audio visually-encoded item reactivates auditory regions as well as visual ones (Wheeler et al., 2000). This highlights a clear link between sensory representations and mnemonic codes. Many studies of human memory have focused on individual senses (for examples, see Weinberger, 2004; Brady et al., 2008; Slotnick et al., 2012; Schurgin, 2018) or chosen to not view memory through a sensory lens at all. However, given that multisensory training has now been shown to benefit learning (Shams and Seitz, 2008), and that episodic memory ties together information across senses in a way that seems to naturally take advantage of crossmodal processing, work in the past two decades has begun to explore the benefits of multisensory stimulus presentation for memory performance.

Research on object recognition has shown that multisensory presentation of objects during the encoding phase seems to enhance later recognition of unisensory representation of the objects. Recognition performance for visual objects presented initially with congruent audio and visual cues was reported to be higher than that of objects initially presented only visually, or with an incongruent audio (Lehmann and Murray, 2005; Thelen et al., 2015). When the recognition test is auditory instead of visual, the pattern of results has been shown to be similar, where multisensory encoding produces higher recognition than audio-alone encoding (Moran et al., 2013).

The aforementioned studies all used a continuous recognition task in which the first and second presentations of the same object are presented within a stream of objects that are interleaved. Experiments that use a more traditional memory paradigm, with distinct encoding and retrieval phases separated by a delay interval, and also those attempting to study more naturalistic tasks have also found a benefit to multisensory encoding. Heikkilä et al. (2015) used such a paradigm to compare benefits in visual recognition to benefits in auditory recognition for stimuli encoded in a multisensory condition compared to stimuli encoded in a unisensory fashion. Contrary to some earlier studies, this study found no benefit to visual recognition between the two conditions, though there was a significant improvement to recognition for auditory memory for items encoded with a visual compared to those encoded as audio only. This study also looked for improvement in recognition of spoken and written words and found that adding audio to written words and vice versa improved recognition, so the benefits seen in previous studies may not be limited to perceptual representations and appear to extend to semantic information. A recent study reported a weak but significant benefit of congruent auditory-visual encoding compared to unisensory or incongruent auditory-visual encoding, in auditory recognition but not in visual recognition (Pecher and Zeelenberg, 2022). In both of these studies, there is an asymmetry in the effect of multisensory encoding on recall: auditory representations benefit from multisensory training whereas visual representations do not. Given that auditory recognition memory is typically noted for being worse than its visual counterpart (Cohen et al., 2009; Gloede and Gregg, 2019), the representations supporting auditory memory may be more ambiguous, and thus may particularly benefit from multisensory encoding.

Findings supporting multisensory benefit to memory performance are not limited to recognition memory paradigms. A recent study showed that recall for visual objects was better when those objects were initially presented with congruent auditory information, even if participants were explicitly told to ignore that auditory information (Duarte et al., 2022). In a similar pattern of results, it was shown that recall of face-name associations could be bolstered by the addition of a name tag that was congruent with the auditory name presentation, extending findings of multisensory memory benefits to associative memory tasks (Murray et al., 2022). These behavioral findings are in line with previous fMRI results showing that higher activation in audiovisual association areas is observed during encoding for face-name pairs that will be later remembered compared with those that will be forgotten (Lee et al., 2017). On the whole, these findings suggest that multisensory encoding is a means by which memory retrieval can be improved, even in complex and naturalistic contexts.

4. Neural mechanisms

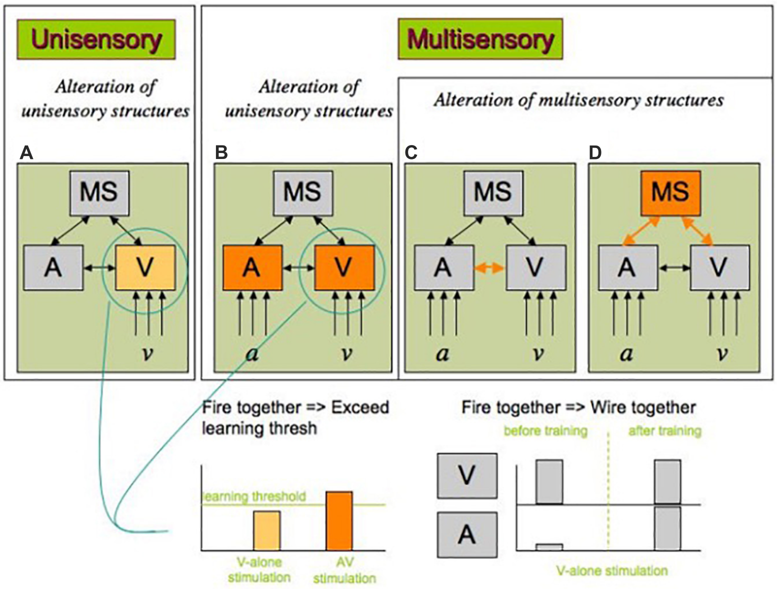
The benefits to perceptual learning, recalibration, adaptation, and memory mentioned thus far have largely been discussed in terms of behavioral studies. This leaves the question of what neural mechanisms may underpin the aforementioned findings and would explain the superiority of multisensory encoding over unisensory encoding/learning. This question remains somewhat open, with many proposed theories holding some weight from the multisensory literature.

Generally, theories fall into two categories: those that make the assumption that learning occurs with neural changes to unisensory regions, and those that make the assumption that learning reflects changes in multisensory structures or crossmodal connectivity (**Figure 2**; Shams and Seitz, 2008). In unisensory theories, the assumption is made that, through training, unisensory regions will eventually refine their processing. This occurs, in a unisensory context, when activity in a unisensory region is heightened above a learning threshold (**Figure 2A**). Under this framework, multisensory training encourages learning by making it easier to elevate the neural activity above the level of the learning threshold, because it activates neural populations both in the sense that is being targeted, and in another region corresponding to another sense that has crossmodal connections to the sense being targeted (**Figure 2B**). These crossmodal connections raise activity in the targeted region above what would be possible if it was stimulated in isolation, making it easier to surpass the learning threshold, and thus leading to faster learning in multisensory training conditions. Such a model could explain the findings that report multisensory encoding of objects does lead to distinct brain activation at retrieval that is not observed with unisensory encoding (Murray et al., 2004; Thelen and Murray, 2013).

By contrast, multisensory frameworks posit that learning is more in line with a Hebbian learning model, following the principle of "fire together, wire together" for the unisensory and multisensory regions (Hebb, 1949; Magee and Grienberger, 2020). Multisensory learning can occur during several different levels under this framework, but we will focus on the idea that plasticity occurs in either the connectivity between unisensory areas that are co-firing during multisensory training (**Figure 2C**) or multisensory regions and their connections to unisensory areas that are strengthened during co-firing (**Figure 2D**). Under either of these mechanisms, learning takes place in part because the two senses contributing to a multisensory signal are co-occurring, which encourages these regions to become more strongly connected. This stronger connection will allow for activation of one region to more easily recruit a larger population of neurons post-training, due to stronger crossmodal connections.

A recent review by Mathias and von Kriegstein (2023), focusing on neuroscience and neurostimulation in the area of multisensory learning came to the conclusion that multisensory mechanisms, consistent with those posited in **Figures 2B–D**, appear to be a better explanation for the observed benefits from multisensory learning as opposed to unisensory learning mechanisms (as would be consistent with those posited in **Figure 2A**). They report on imaging and neurostimulation studies that report that functional connectivity between sensory-specific areas is altered after crossmodal learning [as in von Kriegstein and Giraud (2006), Thelen et al. (2012), Mayer et al. (2015)]. It has also been suggested via simulation studies that both crossmodal connectivity and connections between unisensory regions and higher-level association areas could be strengthened simultaneously during multisensory learning (Cuppini et al., 2017).

However, the aforementioned models of multisensory benefit may not be sufficient to account for some existing phenomena. For example, Barakat et al. (2015) study showed that auditory-only training was able to improve visual rhythm discrimination performance similarly to multisensory training. As there was no stimulation of the visual cortex during training, there was no reason that region should be activated sufficiently to surpass the learning threshold to cause learning as would be expected under unisensory theories (**Figure 2B**). Under multisensory theories, the co-occurrence of the audio and visual signals would be required to change the connectivity between unisensory regions or alter the activation of multisensory regions, and so auditory-only stimulus presentation shouldn't encourage any changes in the visual modality. Barakat et al. (2015) suggest the possibility of a different sort of multisensory activation: one where the crossmodal connections between sensory cortices can be utilized outside of multisensory training (**Figure 3**). Under the assumption that there is pre-existing connectivity between sensory regions



Adapted from Shams & Seitz, *Trends in Cognitive Sciences*, 2008

FIGURE 2
Two possible mechanisms mediating multisensory training advantage for unisensory processing, adapted from Shams and Seitz (2008). **(A)** In classic perceptual learning studies, only one sensory modality (e.g., vision) is trained. In such a unisensory training paradigm, learning would only modify the existing unisensory features (e.g., visual representations, *v,* or auditory representations, *a*, here). In multisensory training paradigms **(B–D)** multiple sensory modalities (e.g., vision and hearing) are stimulated simultaneously. The advantage of multisensory training over unisensory training could be due to **(B)** the fact that the pre-existing connection between the sensory regions (A and V, here) gives rise to a higher activity of each unisensory region (e.g., V) as compared to unisensory stimulation and exceeds the threshold required for learning to occur. Alternatively, multisensory training which involves repeated co-activation of unisensory regions A and V could result in strengthening of multisensory structures (MS here), such as direct connection between unisensory regions, as depicted in **(C)** or the connection between unisensory regions and multisensory regions, as depicted in **(D)**, or both in a "fire together, wire together" fashion. As a result of this new wiring, the activation of one unisensory region can lead to activation of the other unisensory region [either via direct connection **(C)** or indirectly through multisensory connections **(D)** or both], in effect implementing redintegration (see section "4. Neural mechanisms" for more detail).

(e.g., Eckert et al., 2008; Beer et al., 2011) and also between sensory regions and decision regions (e.g., Heekeren et al., 2008; Siegel et al., 2011), this could operate in two ways. It is possible that one sensory region could "teach" another– in the example of Barakat et al. (2015), the auditory region is able to "teach" the visual region (**Figure 3A**). At test, the visual region is activated, and this will, in turn, cause partial activation of the auditory region, due to their crossmodal connections. Training of the participant in the auditory condition will result in refined processing within the auditory cortex, and activation of this region will allow for signals from auditory cortical regions to help refine the visual processing, improving visual performance. Alternatively, due to crossmodal connections and the putative superiority of the auditory cortex in temporal processing (Glenberg et al., 1989; Repp and Penel, 2002; Mcauley and Henry, 2010; Grahn, 2012), the visual region could outsource processing on this task to the auditory region almost entirely (**Figure 3B**). Here, activation of the visual region would excite the auditory region through crossmodal pathways and, as the trained auditory region is thus activated sufficiently to be used in the decision-making process for the visual decision. Under either of these models, it is possible for unisensory training in one modality to influence performance in another modality, provided the regions are connected crossmodally or via a multisensory convergence area. Still, it is not clear why multisensory training would not result in a superior outcome to auditory-alone training. Future studies will need to explore the role of relative dominance of the two modalities in a given task as well as other factors such as task difficulty and duration of training to shed light on the underlying mechanisms and the factors that determine the effectiveness of multisensory training relative to unisensory training in a given task for a given individual.

5. Discussion and future directions

In the realms of human learning and memory, it has been continually shown that taking advantage of multisensory training/encoding can improve later performance, including performance in unisensory tasks. Exposure to correlated or redundant crossmodal stimuli has been shown to lead to faster learning and enhanced unisensory processing in perceptual learning tasks (as in Kim et al., 2008; Barakat et al., 2015). Similarly, passive exposure to co-occurring sensory input across modalities (resulting in the acquisition of a novel association) can also lead to improved unisensory processing (as in Ernst, 2007; Seitz et al., 2007). Repeated mismatch across the senses can also result in learning via recalibration of sensory representations (as in Wozny and Shams, 2011). Multisensory encoding of stimuli has been shown to improve later recall for visual and auditory stimuli, even when recall cues are unisensory (as in Lehmann and Murray, 2005; Moran et al., 2013; Duarte et al., 2022; Murray et al., 2022). Altogether these results clearly show that the human nervous

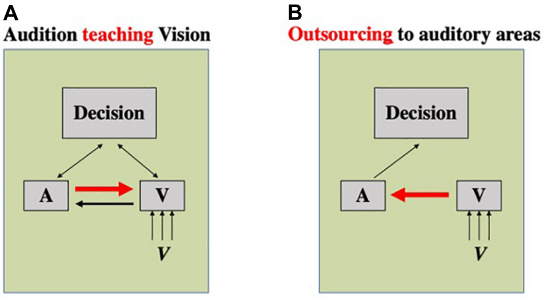


FIGURE 3
Two possible mechanisms underlying crossmodal transfer of perceptual learning. Boxes A, V, and decision represent, respectively, auditory and visual processing stages, and a decision-making stage of processing. **(A)** A mechanism wherein one sense is able to teach another. In this case, a sense that is superior in performing a task (here, auditory modality) would be able to teach that information to a different sense (here, vision) through crossmodal connections. **(B)** A mechanism wherein one sense outsources the processing to another sense. In this case, a sense that is worse in performing a task (here, vision) will send information to a sensory region that is more apt in processing that task (here, audition) through crossmodal connections.

system is acutely sensitive to the relationship between sensory signals across modalities, and exposure to multisensory stimuli, not only refines multisensory processing (see Quintero et al., 2022; Mathias and von Kriegstein, 2023; for reviews), but it also alters and refines unisensory representations and the ensuing unisensory processing.

While we have posited possible models for the observed improvement above, it should be noted that this is non-exhaustive–several possible mechanisms may be at play, separately or in combination. While Mathias and von Kriegstein (2023) point out that multisensory models capture neuroscientific evidence better, many important questions regarding the neural mechanisms of perceptual learning remain unanswered. For example, it is not clear to what degree and under which conditions the benefits of multisensory training and encoding stem from alterations in crossmodal connectivity versus changes in activity of multisensory regions versus refined representations in unisensory regions. Some recent work in animals even suggests that multimodal experience fundamentally changes the cooperative nature of how senses relate; they claim that the natural interaction of the senses is one of competition, which can be shaped into cooperation through multisensory experience (Yu et al., 2019; Wang et al., 2020). If such cooperative organization is truly only available with multisensory experience, then multisensory learning may reflect an even more complex shift in the relationship between multimodal and unisensory brain regions. It is also not clear under which conditions "unisensory" processing regions (such as visual cortex or auditory cortex) are involved in providing a "teaching signal" to another modality and/or outsource processing to another sensory region. Clarifying which circuits or pathways best capture learning and memory benefits stemming from multisensory exposure should be the focus of future research. Understanding these neural mechanisms would allow us to better understand and harness them for improving human learning and memory performance.

Perhaps an even more important target for further research would be to uncover computational principles governing multisensory learning. While some general ideas have been proposed in the literature there are few attempts to comprehensively and rigorously model how the brain benefits from multisensory stimulus presentation in learning/memory contexts. Rigorous computational modeling is needed to shed light on the nature of information processing involved in the different sensory conditions during learning and provide an understanding of how it is possible to achieve the same level of accuracy in unisensory conditions and multisensory conditions after multisensory training (see the discussion of Kim et al., 2008 in the section "2.1. Perceptual learning").

With regards to memory, there are many behavioral observations that span decades supporting that multisensory and unisensory information appears to interact in the memory system, yet computational models are lacking. For example, the phenomenon of redintegration (Horowitz and Prytulak, 1969), where unisensory information can cue a memory with information across multiple senses, has been long cited as a behavioral phenomenon, yet the mechanism by which the senses are entangled in memory remain unclear. Mathias and von Kriegstein (2023) review computational approaches to this question and propose that a Predictive Coding framework can account for some of the findings. While this is a good start, future studies should engage in model comparison and aim to offer computational models that can quantitatively account for the empirical findings. Computational models are needed to formalize an understanding of the way sensory cues work in memory, and to make testable predictions about conditions and the nature of crossmodal interactions and presence and type of multisensory benefit in learning across tasks and sensory conditions.

A better mechanistic and computational understanding of the mechanisms behind multisensory learning and memory benefits would also allow for us to better harness these mechanisms and principles to improve memory and learning in everyday life. Multisensory stimulus presentation is often relatively simple to implement, especially with current technologies, and would provide an easy avenue to bolster learning and memory in a number of contexts. As discussed, the above studies of implicit learning have shown that even arbitrary associations can be quickly learned, and subsequently serve as the basis for improved unisensory processing. Therefore, the benefits of multisensory training/encoding are not limited to only naturalistic tasks. Further research into how multisensory benefits could be applied to everyday tasks could provide a useful avenue to improve human cognitive performance in day-to-day life and guide the development of more effective educational and clinical practice.

The recent findings on benefits of multisensory learning as reviewed here and elsewhere (Shams and Seitz, 2008; Mathias and von Kriegstein, 2023) are also noteworthy in that they may warrant a shift in how the fields of neuroscience and psychology view perceptual learning. These findings have generally been framed [including by us in Shams and Seitz (2008)] as superiority of multisensory learning over unisensory learning. However, a more rational framing may be to view them as showing the inferiority of unisensory learning compared to multisensory learning. In other words, it can be argued that the longstanding tradition of studying learning in unisensory settings has biased interpretation of these findings as reflecting a multisensory benefit, as opposed to recognizing a disadvantage in the unisensory protocols. The world around us provides constant crossmodal information– it's possible the brain would develop to treat this as a "default" level of information available for learning and memory. If the brain is truly developed to utilize multisensory cues when learning about the environment, then providing less information, as in unisensory learning paradigms, could be forcing the system to use impoverished computational resources for learning. This would lead to an inferior outcome for learning compared to when multisensory cues are available and full computational resources would be used. Under this assumption, multisensory perception is the naturalistic baseline for the brain, which unisensory approaches cannot fully explore. Just as we need information across many senses to truly understand our world, we will need to study the dynamic interplay between the senses to truly understand the human mind.

Author contributions

LS and CM: conceptualization and writing. Both authors contributed to the article and approved the submitted version.

Acknowledgments

We thank Aaron Seitz for his help with locating the data used for completing **Figure 1B**.

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

References

Adeli, M., Rouat, J., and Molotchnikoff, S. (2014). Audiovisual correspondence between musical timbre and visual shapes. *Front. Hum. Neurosci.* 8:352. doi: 10.3389/fnhum.2014.00352

Alais, D., and Cass, J. (2010). Multisensory perceptual learning of temporal order: Audiovisual learning transfers to vision but not audition. *PLoS One* 5:e11283. doi: 10.1371/journal.pone.0011283

Aller, M., and Noppeney, U. (2019). To integrate or not to integrate: Temporal dynamics of hierarchical Bayesian causal inference. *PLoS Biol.* 17:e3000210. doi: 10.1371/journal.pbio.3000210

Aller, M., Mihalik, A., and Noppeney, U. (2022). Audiovisual adaptation is expressed in spatial and decisional codes. *Nat. Commun.* 13:1. doi: 10.1038/s41467-022-31549-0

Aslin, R. N. (2017). Statistical learning: A powerful mechanism that operates by mere exposure. *Wiley Interdiscip. Rev. Cogn. Sci.* 8:e1373. doi: 10.1002/wcs.1373

Barakat, B., Seitz, A. R., and Shams, L. (2015). Visual rhythm perception improves through auditory but not visual training. *Curr. Biol.* 25, R60–R61. doi: 10.1016/j.cub.2014.12.011

Batterink, L. J., Paller, K. A., and Reber, P. J. (2019). Understanding the neural bases of implicit and statistical learning. *Top. Cogn. Sci.* 11, 482–503. doi: 10.1111/tops.12420

Beer, A. L., Plank, T., and Greenlee, M. W. (2011). Diffusion tensor imaging shows white matter tracts between human auditory and visual cortex. *Exp. Brain Res.* 213, 299–308. doi: 10.1007/s00221-011-2715-y

Botvinick, M., and Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature* 391:6669. doi: 10.1038/35784

Brady, T. F., Konkle, T., Alvarez, G. A., and Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proc. Natl. Acad. Sci. U.S.A.* 105, 14325–14329. doi: 10.1073/pnas.0803390105

Brunel, L., Carvalho, P. F., and Goldstone, R. L. (2015). It does belong together: Cross-modal correspondences influence cross-modal integration during perceptual learning. *Front. Psychol.* 6:358. doi: 10.3389/fpsyg.2015.00358

Bruns, P. (2019). The ventriloquist illusion as a tool to study multisensory processing: An update. *Front. Integr. Neurosci.* 13:51. doi: 10.3389/fnint.2019.00051

Burr, D., and Gori, M. (2012). "Multisensory integration develops late in humans," in *The neural bases of multisensory processes*, eds M. M. Murray and M. T. Wallace (Boca Raton, FL: CRC Press/Taylor and Francis).

Calford, M. B. (2002). Dynamic representational plasticity in sensory cortex. *Neuroscience* 111, 709–738. doi: 10.1016/S0306-4522(02)00022-2

Cao, Y., Summerfield, C., Park, H., Giordano, B. L., and Kayser, C. (2019). Causal inference in the multisensory brain. *Neuron* 102, 1076–1087.e8. doi: 10.1016/j.neuron.2019.03.043

Cohen, M. A., Horowitz, T. S., and Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6008–6010. doi: 10.1073/pnas.0811884106

Conway, C. M., and Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 24–39. doi: 10.1037/0278-7393.31.1.24

Conway, C. M., and Christiansen, M. H. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. *Psychol. Sci.* 17, 905–912. doi: 10.1111/j.1467-9280.2006.01801.x

Cuppini, C., Ursino, M., Magosso, E., Ross, L. A., Foxe, J. J., and Molholm, S. (2017). A computational analysis of neural mechanisms underlying the maturation of multisensory speech integration in neurotypical children and those on the autism spectrum. *Front. Hum. Neurosci.* 11:518. doi: 10.3389/fnhum.2017.00518

De Niear, M. A., Gupta, P. B., Baum, S. H., and Wallace, M. T. (2018). Perceptual training enhances temporal acuity for multisensory speech. *Neurobiol. Learn. Mem.* 147, 9–17. doi: 10.1016/j.nlm.2017.10.016

Dionne-Dostie, E., Paquette, N., Lassonde, M., and Gallagher, A. (2015). Multisensory integration and child neurodevelopment. *Brain Sci.* 5, 32–57. doi: 10.3390/brainsci5010032

Driver, J., and Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory-specific', brain regions, neural responses, and judgments. *Neuron* 57, 11–23. doi: 10.1016/j.neuron.2007.12.013

Duarte, S. E., Ghetti, S., and Geng, J. J. (2022). Object memory is multisensory: Task-irrelevant sounds improve recollection. *Psychon. Bull. Rev.* 30, 652–665. doi: 10.3758/s13423-022-02182-1

Eckert, M. A., Kamdar, N. V., Chang, C. E., Beckmann, C. F., Greicius, M. D., and Menon, V. (2008). A cross-modal system linking primary auditory and visual cortices: Evidence from intrinsic fMRI connectivity analysis. *Hum. Brain Mapp.* 29, 848–857. doi: 10.1002/hbm.20560

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *J. Vis.* 7, 7.1–14. doi: 10.1167/7.5.7

Ferrari, A., and Noppeney, U. (2021). Attention controls multisensory perception via two distinct mechanisms at different levels of the cortical hierarchy. *PLoS Biol.* 19:e3001465. doi: 10.1371/journal.pbio.3001465

Ferraro, S., Van Ackeren, M. J., Mai, R., Tassi, L., Cardinale, F., Nigri, A., et al. (2020). Stereotactic electroencephalography in humans reveals multisensory signal in early visual and auditory cortices. *Cortex* 126, 253–264. doi: 10.1016/j.cortex.2019.12.032

Fiser, J., and Lengyel, G. (2022). Statistical learning in vision. *Annu. Rev. Vis. Sci.* 8, 265–290. doi: 10.1146/annurev-vision-100720-103343

Frost, R., Armstrong, B. C., Siegelman, N., and Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends Cogn. Sci.* 19, 117–125. doi: 10.1016/j.tics.2014.12.010

Fujisaki, W., Shimojo, S., Kashino, M., and Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nat. Neurosci.* 7:7. doi: 10.1038/nn1268

Gau, R., Bazin, P.-L., Trampel, R., Turner, R., and Noppeney, U. (2020). Resolving multisensory and attentional influences across cortical depth in sensory cortices. *ELife* 9:e46856. doi: 10.7554/eLife.46856

Ghazanfar, A., and Schroeder, C. (2006). Is neocortex essentially multisensory? *Trends Cogn. Sci.* 10, 278–285. doi: 10.1016/j.tics.2006.04.008

Glenberg, A. M., Mann, S., Altman, L., Forman, T., and Procise, S. (1989). Modality effects in the coding reproduction of rhythms. *Mem. Cogn.* 17, 373–383. doi: 10.3758/BF03202611

Gloede, M. E., and Gregg, M. K. (2019). The fidelity of visual and auditory memory. *Psychono. Bull. Rev.* 26, 1325–1332. doi: 10.3758/s13423-019-01597-7

Gold, J. I., and Watanabe, T. (2010). Perceptual learning. *Curr. Biol.* 20, R46–R48. doi: 10.1016/j.cub.2009.10.066

Goldstone, R. L. (1998). Perceptual learning. *Annu. Rev. Psychol.* 49, 585–612. doi: 10.1146/annurev.psych.49.1.585

Gori, M., Del Viva, M., Sandini, G., and Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Curr. Biol.* 18, 694–698. doi: 10.1016/j.cub.2008.04.036

Grahn, J. A. (2012). See what I hear? Beat perception in auditory and visual rhythms. *Exp. Brain Res.* 220, 51–61. doi: 10.1007/s00221-012-3114-8

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. Hoboken, NJ: Wiley, 335.

Heekeren, H. R., Marrett, S., and Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. *Nat. Rev. Neurosci.* 9:6. doi: 10.1038/nrn2374

Heikkilä, J., Alho, K., Hyvönen, H., and Tiippana, K. (2015). Audiovisual semantic congruency during encoding enhances memory performance. *Exp. Psychol.* 62, 123–130. doi: 10.1027/1618-3169/a000279

Held, R., Ostrovsky, Y., de Gelder, B., Gandhi, T., Ganesh, S., Mathur, U., et al. (2011). The newly sighted fail to match seen with felt. *Nat. Neurosci.* 14:5. doi: 10.1038/nn.2795

Hirst, R. J., McGovern, D. P., Setti, A., Shams, L., and Newell, F. N. (2020). What you see is what you hear: Twenty years of research using the Sound-Induced Flash Illusion. *Neurosci. Biobehav. Rev.* 118, 759–774. doi: 10.1016/j.neubiorev.2020.09.006

Hong, F., Badde, S., and Landy, M. S. (2021). Causal inference regulates audiovisual spatial recalibration via its influence on audiovisual perception. *PLoS Comput. Biol.* 17:e1008877. doi: 10.1371/journal.pcbi.1008877

Horowitz, L. M., and Prytulak, L. S. (1969). Redintegrative memory. *Psychol. Rev.* 76, 519–531. doi: 10.1037/h0028139

Horsfall, R. P., Wuerger, S. M., and Meyer, G. F. (2021). Narrowing of the audiovisual temporal binding window due to perceptual training is specific to high visual intensity stimuli. *I-Perception* 12:204166952097867. doi: 10.1177/2041669520978670

Jraissati, Y., Slobodenyuk, N., Kanso, A., Ghanem, L., and Elhajj, I. (2016). Haptic and tactile adjectives are consistently mapped onto color space. *Multisens. Res.* 29, 253–278. doi: 10.1163/22134808-00002512

Kayser, C., and Shams, L. (2015). Multisensory causal inference in the brain. *PLoS Biol.* 13:e1002075. doi: 10.1371/journal.pbio.1002075

Kim, R. S., Seitz, A. R., and Shams, L. (2008). Benefits of stimulus congruency for multisensory facilitation of visual learning. *PLoS One* 3:e1532. doi: 10.1371/journal.pone.0001532

Knowlton, B. J., Squire, L. R., and Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learn. Mem.* 1, 106–120. doi: 10.1101/lm.1.2.106

Knudsen, E. (1994). Supervised learning in the brain. *J. Neurosci.* 14, 3985–3997. doi: 10.1523/JNEUROSCI.14-07-03985.1994

Knudsen, E. I., and Knudsen, P. F. (1985). Vision guides the adjustment of auditory localization in young barn owls. *Science* 230, 545–548. doi: 10.1126/science.4048948

Kolb, B., and Whishaw, I. Q. (1998). Brain plasticity and behavior. *Annu. Rev. Psychol.* 49, 43–64. doi: 10.1146/annurev.psych.49.1.43

Kopco, N., Lin, I.-F., Shinn-Cunningham, B. G., and Groh, J. M. (2009). Reference frame of the ventriloquism aftereffect. *J. Neurosci.* 29, 13809–13814. doi: 10.1523/JNEUROSCI.2783-09.2009

Kramer, A., Röder, B., and Bruns, P. (2020). Feedback modulates audio-visual spatial recalibration. *Front. Integr. Neurosci.* 13:74. doi: 10.3389/fnint.2019.00074

Landry, S. P., Guillemot, J.-P., and Champoux, F. (2013). Temporary deafness can impair multisensory integration: A study of cochlear-implant users. *Psychol. Sci.* 24, 1260–1268. doi: 10.1177/0956797612471142

Lee, H., Stirnberg, R., Stöcker, T., and Axmacher, N. (2017). Audiovisual integration supports face–name associative memory formation. *Cogn. Neurosci.* 8, 177–192. doi: 10.1080/17588928.2017.1327426

Lehmann, S., and Murray, M. M. (2005). The role of multisensory memories in unisensory object discrimination. *Cogn. Brain Res.* 24, 326–334. doi: 10.1016/j.cogbrainres.2005.02.005

Magee, J. C., and Grienberger, C. (2020). Synaptic plasticity forms and functions. *Annu. Rev. Neurosci.* 43, 95–117. doi: 10.1146/annurev-neuro-090919-022842

Mathias, B., and von Kriegstein, K. (2023). Enriched learning: Behavior, brain, and computation. *Trends Cogn. Sci.* 27, 81–97. doi: 10.1016/j.tics.2022.10.007

Mayer, K. M., Yildiz, I. B., Macedonia, M., and von Kriegstein, K. (2015). Visual and motor cortices differentially support the translation of foreign language words. *Curr. Biol.* 25, 530–535. doi: 10.1016/j.cub.2014.11.068

Mayer, R. E. (2014). *The cambridge handbook of multimedia learning*. Cambridge: Cambridge University Press.

Mcauley, J. D., and Henry, M. J. (2010). Modality effects in rhythm processing: Auditory encoding of visual rhythms is neither obligatory nor automatic. *Attent. Percept. Psychophys.* 72, 1377–1389. doi: 10.3758/APP.72.5.1377

McGovern, D. P., Astle, A. T., Clavin, S. L., and Newell, F. N. (2016). Task-specific transfer of perceptual learning across sensory modalities. *Curr. Biol.* 26, R20–R21. doi: 10.1016/j.cub.2015.11.048

McGovern, D. P., Burns, S., Hirst, R. J., and Newell, F. N. (2022). Perceptual training narrows the temporal binding window of audiovisual integration in both younger and older adults. *Neuropsychologia* 173:108309. doi: 10.1016/j.neuropsychologia.2022.108309

Montessori, M. (2013). *The montessori method*. Piscataway, NJ: Transaction Publishers.

Moran, Z. D., Bachman, P., Pham, P., Hah Cho, S., Cannon, T. D., and Shams, L. (2013). Multisensory encoding improves auditory recognition. *Multisens. Res.* 26, 581–592. doi: 10.1163/22134808-00002436

Murray, C. A., Tarlow, M., Rissman, J., and Shams, L. (2022). Multisensory encoding of names via name tags facilitates remembering. *Appl. Cogn. Psychol.* 36, 1277–1291. doi: 10.1002/acp.4012

Murray, M. M., Lewkowicz, D. J., Amedi, A., and Wallace, M. T. (2016a). Multisensory processes: A balancing act across the lifespan. *Trends Neurosci.* 39, 567–579. doi: 10.1016/j.tins.2016.05.003

Murray, M. M., Michel, C. M., Grave de Peralta, R., Ortigue, S., Brunet, D., Gonzalez Andino, S., et al. (2004). Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *NeuroImage* 21, 125–135. doi: 10.1016/j.neuroimage.2003.09.035

Murray, M. M., Thelen, A., Thut, G., Romei, V., Martuzzi, R., and Matusz, P. J. (2016b). The multisensory function of the human primary visual cortex. *Neuropsychologia* 83, 161–169. doi: 10.1016/j.neuropsychologia.2015.08.011

Nardini, M., and Cowie, D. (2012). "The development of multisensory balance, locomotion, orientation, and navigation," in *Multisensory development*, eds A. J. Bremner, D. J. Lewkowicz, and C. Spence (Oxford: Oxford University Press), 137–158. doi: 10.1093/acprof:oso/9780199586059.003.0006

Nyberg, L., Habib, R., McIntosh, A. R., and Tulving, E. (2000). Reactivation of encoding-related brain activity during memory retrieval. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11120–11124. doi: 10.1073/pnas.97.20.11120

O'Brien, J. M., Chan, J. S., and Setti, A. (2020). Audio-visual training in older adults: 2-interval-forced choice task improves performance. *Front. Neurosci.* 14:569212. doi: 10.3389/fnins.2020.569212

O'Brien, J., Mason, A., Chan, J., and Setti, A. (2023). Can we train multisensory integration in adults? A systematic review. *Multisens. Res.* 36, 111–180. doi: 10.1163/22134808-bja10090

Parise, C. V. (2016). Crossmodal correspondences: Standing issues and experimental guidelines. *Multisens. Res.* 29, 7–28. doi: 10.1163/22134808-00002502

Parise, C. V., and Spence, C. (2009). 'When birds of a feather flock together': Synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS One* 4:e5664. doi: 10.1371/journal.pone.0005664

Pecher, D., and Zeelenberg, R. (2022). Does multisensory study benefit memory for pictures and sounds? *Cognition* 226:105181. doi: 10.1016/j.cognition.2022.105181

Peters, M. A. K., Balzer, J., and Shams, L. (2015). Smaller = denser, and the brain knows it: Natural statistics of object density shape weight expectations. *PLoS One* 10:e0119794. doi: 10.1371/journal.pone.0119794

Powers, A. R., Hillock-Dunn, A., and Wallace, M. T. (2016). Generalization of multisensory perceptual learning. *Sci. Rep.* 6:23374. doi: 10.1038/srep23374

Powers, A. R. III, Hillock, A. R., and Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *J. Neurosci.* 29, 12265–12274. doi: 10.1523/JNEUROSCI.3501-09.2009

Quintero, S. I., Shams, L., and Kamal, K. (2022). Changing the tendency to integrate the senses. *Brain Sci.* 12:10. doi: 10.3390/brainsci12101384

Reber, A. S. (1967). Implicit learning of artificial grammars. *J. Verbal Learn. Verbal Behav.* 6, 855–863. doi: 10.1016/S0022-5371(67)80149-X

Recanzone, G. H. (1998). Rapidly induced auditory plasticity: The ventriloquism?aftereffect. *Proc. Natl. Acad. Sci. U.S.A.* 95, 869–875. doi: 10.1073/pnas.95.3.869

Repp, B. H., and Penel, A. (2002). Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 1085–1099. doi: 10.1037/0096-1523.28.5.1085

Rohe, T., and Noppeney, U. (2015). Cortical hierarchies perform bayesian causal inference in multisensory perception. *PLoS Biol.* 13:e1002073. doi: 10.1371/journal.pbio.1002073

Rohe, T., Ehlis, A.-C., and Noppeney, U. (2019). The neural dynamics of hierarchical Bayesian causal inference in multisensory perception. *Nat. Commun.* 10:1. doi: 10.1038/s41467-019-09664-2

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928. doi: 10.1126/science.274.5294.1926

Schurgin, M. W. (2018). Visual memory, the long and the short of it: A review of visual working memory and long-term memory. *Attent. Percept. Psychophys.* 80, 1035–1056. doi: 10.3758/s13414-018-1522-y

Seitz, A. R., Kim, R., and Shams, L. (2006). Sound facilitates visual learning. *Curr. Biol.* 16, 1422–1427. doi: 10.1016/j.cub.2006.05.048

Seitz, A. R., Kim, R., van Wassenhove, V., and Shams, L. (2007). Simultaneous and independent acquisition of multisensory and unisensory associations. *Perception* 36, 1445–1453. doi: 10.1068/p5843

Setti, A., Finnigan, S., Sobolewski, R., McLaren, L., Robertson, I. H., Reilly, R. B., et al. (2011). Audiovisual temporal discrimination is less efficient with aging: An event-related potential study. *NeuroReport* 22:554. doi: 10.1097/WNR.0b013e328348c731

Setti, A., Stapleton, J., Leahy, D., Walsh, C., Kenny, R. A., and Newell, F. N. (2014). Improving the efficiency of multisensory integration in older adults: Audio-visual temporal discrimination training reduces susceptibility to the sound-induced flash illusion. *Neuropsychologia* 61, 259–268. doi: 10.1016/j.neuropsychologia.2014.06.027

Shams, L., and Seitz, A. R. (2008). Benefits of multisensory learning. *Trends Cogn. Sci.* 12, 411–417. doi: 10.1016/j.tics.2008.07.006

Shams, L., Kamitani, Y., and Shimojo, S. (2000). What you see is what you hear. *Nature* 408:6814. doi: 10.1038/35048669

Shams, L., Wozny, D. R., Kim, R. S., and Seitz, A. (2011). Influences of multisensory experience on subsequent unisensory processing. *Front. Psychol.* 2:264. doi: 10.3389/fpsyg.2011.00264

Sherman, B. E., Graves, K. N., and Turk-Browne, N. B. (2020). The prevalence and importance of statistical learning in human cognition and behavior. *Curr. Opin. Behav. Sci.* 32, 15–20. doi: 10.1016/j.cobeha.2020.01.015

Siegel, M., Engel, A., and Donner, T. (2011). Cortical network dynamics of perceptual decision-making in the human brain. *Front. Hum. Neurosci.* 5:21. doi: 10.3389/fnhum.2011.00021

Slotnick, S. D., Thompson, W. L., and Kosslyn, S. M. (2012). Visual memory and visual mental imagery recruit common control and sensory regions of the brain. *Cogn. Neurosci.* 3, 14–20. doi: 10.1080/17588928.2011.578210

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attent. Percept. Psychophys.* 73, 971–995. doi: 10.3758/s13414-010-0073-7

Spence, C., Levitan, C. A., Shankar, M. U., and Zampini, M. (2010). Does food color influence taste and flavor perception in humans? *Chemosens. Percept.* 3, 68–84. doi: 10.1007/s12078-010-9067-z

Stevenson, R., Sheffield, S. W., Butera, I. M., Gifford, R. H., and Wallace, M. (2017). Multisensory integration in cochlear implant recipients. *Ear Hear.* 38, 521–538. doi: 10.1097/AUD.0000000000000435

Sürig, R., Bottari, D., and Röder, B. (2018). Transfer of audio-visual temporal training to temporal and spatial audio-visual tasks. *Multisens. Res.* 31, 556–578. doi: 10.1163/22134808-00002611

Thelen, A., and Murray, M. M. (2013). The efficacy of single-trial multisensory memories. *Multisens. Res.* 26, 483–502. doi: 10.1163/22134808-00002426

Thelen, A., Cappe, C., and Murray, M. M. (2012). Electrical neuroimaging of memory discrimination based on single-trial multisensory learning. *NeuroImage* 62, 1478–1488. doi: 10.1016/j.neuroimage.2012.05.027

Thelen, A., Talsma, D., and Murray, M. M. (2015). Single-trial multisensory memories affect later auditory and visual object discrimination. *Cognition* 138, 148–160. doi: 10.1016/j.cognition.2015.02.003

Thurlow, W. R., and Jack, C. E. (1973). Certain Determinants of the "Ventriloquism Effect." *Percept. Mot. Skills* 36(3_suppl), 1171–1184. doi: 10.2466/pms.1973.36.3c.1171

Tulving, E. (1993). What is episodic memory? *Curr. Direct. Psychol. Sci.* 2, 67–70.

Virsu, V., Oksanen-Hennah, H., Vedenpää, A., Jaatinen, P., and Lahti-Nuuttila, P. (2008). Simultaneity learning in vision, audition, tactile sense and their cross-modal combinations. *Exp. Brain Res.* 186, 525–537. doi: 10.1007/s00221-007-1254-z

von Kriegstein, K., and Giraud, A.-L. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biol.* 4:e326. doi: 10.1371/journal.pbio.0040326

Vroomen, J., Keetels, M., de Gelder, B., and Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cogn. Brain Res.* 22, 32–35. doi: 10.1016/j.cogbrainres.2004.07.003

Walk, A. M., and Conway, C. M. (2016). Cross-domain statistical–sequential dependencies are difficult to learn. *Front. Psychol.* 7:250. doi: 10.3389/fpsyg.2016.00250

Walker, P., Scallon, G., and Francis, B. (2017). Cross-sensory correspondences: Heaviness is dark and low-pitched. *Perception* 46:7. doi: 10.1177/0301006616684369

Wallace, M. T., Meredith, M. A., and Stein, B. E. (1998). Multisensory integration in the superior colliculus of the alert cat. *J. Neurophysiol.* 80, 1006–1010. doi: 10.1152/jn.1998.80.2.1006

Wallace, M. T., Perrault, T. J., Hairston, W. D., and Stein, B. E. (2004). Visual experience is necessary for the development of multisensory integration. *J. Neurosci.* 24, 9580–9584. doi: 10.1523/JNEUROSCI.2535-04.2004

Wang, Z., Yu, L., Xu, J., Stein, B. E., and Rowland, B. A. (2020). Experience creates the multisensory transform in the superior colliculus. *Front. Integr. Neurosci.* 14:18. doi: 10.3389/fnint.2020.00018

Weinberger, N. M. (2004). Specific long-term memory traces in primary auditory cortex. *Nat. Rev. Neurosci.* 5:4. doi: 10.1038/nrn1366

Wheeler, M. E., Petersen, S. E., and Buckner, R. L. (2000). Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11125–11129. doi: 10.1073/pnas.97.20.11125

Wozny, D. R., Beierholm, U. R., and Shams, L. (2008). Human trimodal perception follows optimal statistical inference. *J. Vis.* 8:24. doi: 10.1167/8.3.24

Wozny, D., and Shams, L. (2011). Computational characterization of visually induced auditory spatial adaptation. *Front. Integr. Neurosci.* 5:75. doi: 10.3389/fnint.2011.00075

Xu, J., Yu, L., Rowland, B. A., Stanford, T. R., and Stein, B. E. (2014). Noise-rearing disrupts the maturation of multisensory integration. *Eur. J. Neurosci.* 39, 602–613. doi: 10.1111/ejn.12423

Yu, L., Cuppini, C., Xu, J., Rowland, B. A., and Stein, B. E. (2019). Cross-modal competition: The default computation for multisensory processing. *J. Neurosci.* 39, 1374–1385. doi: 10.1523/JNEUROSCI.1806-18.2018

Zwiers, M. P., Van Opstal, A. J., and Paige, G. D. (2003). Plasticity in human sound localization induced by compressed spatial vision. *Nat. Neurosci.* 6:2. doi: 10.1038/nn999

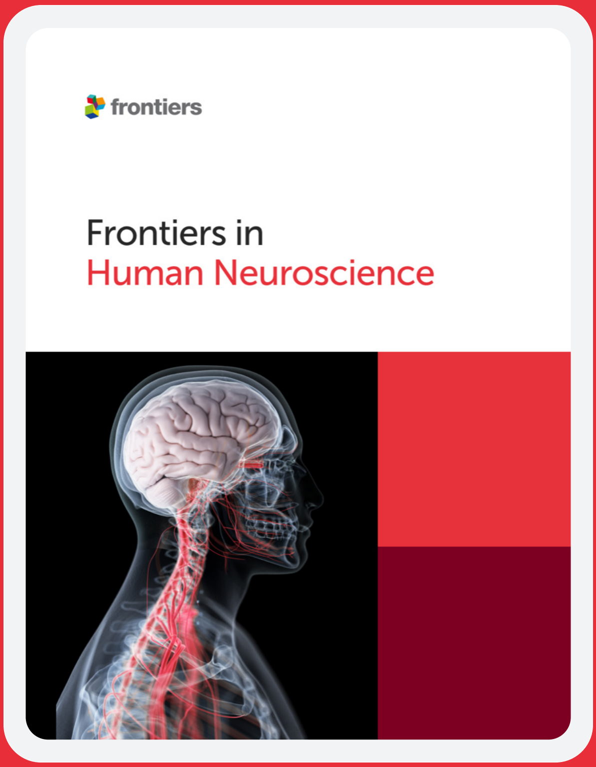
Frontiers in
Human Neuroscience

Bridges neuroscience and psychology to understand the human brain

The second most-cited journal in the field of psychology, that bridges research in psychology and neuroscience to advance our understanding of the human brain in both healthy and diseased states.

Discover the latest Research Topics

See more →



frontiers

Frontiers in
Human Neuroscience

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact