NeuroImage 285 (2024) 120483

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Enhancing learning outcomes through multisensory integration: A fMRI study of audio-visual training in virtual reality

Kholoud Alwashmi [a,d,*], Georg Meyer [b], Fiona Rowe [c], Ryan Ward [b,e]

[a] Faculty of Health and Life Sciences, University of Liverpool, United Kingdom
[b] Digital Innovation Facility, University of Liverpool, United Kingdom
[c] Institute of Population Health, University of Liverpool, United Kingdom
[d] Department of Radiology, Princess Nourah bint Abdulrahman University, Saudi Arabia
[e] School Computer Science and Mathematics, Liverpool John Moores University, United Kingdom

ARTICLE INFO

Keywords:
fMRI
Multisensory
Audio-visual
Learning
Virtual-reality
Eye-movement

ABSTRACT

The integration of information from different sensory modalities is a fundamental process that enhances perception and performance in real and virtual environments (VR). Understanding these mechanisms, especially during learning tasks that exploit novel multisensory cue combinations provides opportunities for the development of new rehabilitative interventions.

This study aimed to investigate how functional brain changes support behavioural performance improvements during an audio-visual (AV) learning task. Twenty healthy participants underwent a 30 min daily VR training for four weeks. The task was an AV adaptation of a 'scanning training' paradigm that is commonly used in hemianopia rehabilitation. Functional magnetic resonance imaging (fMRI) and performance data were collected at baseline, after two and four weeks of training, and four weeks post-training.

We show that behavioural performance, operationalised as mean reaction time reduction in VR, significantly improves. In separate tests in a controlled laboratory environment, we showed that the behavioural performance gains in the VR training environment transferred to a significant mean RT reduction for the trained AV voluntary task on a computer screen. Enhancements were observed in both the visual-only and AV conditions, with the latter demonstrating a faster response time supported by the presence of audio cues. The behavioural learning effect also transfers to two additional tasks that were tested: a visual search task and an involuntary visual task.

Our fMRI results reveal an increase in functional activation (BOLD signal) in multisensory brain regions involved in early-stage AV processing: the thalamus, the caudal inferior parietal lobe and cerebellum. These functional changes were only observed for the trained, multisensory, task and not for unimodal visual stimulation. Functional activation changes in the thalamus were significantly correlated to behavioural performance improvements.

This study demonstrates that incorporating spatial auditory cues to voluntary visual training in VR leads to augmented brain activation changes in multisensory integration, resulting in measurable performance gains across tasks. The findings highlight the potential of VR-based multisensory training as an effective method for enhancing cognitive function and as a potentially valuable tool in rehabilitative programmes.

Abbreviations: VR, virtual reality; AV, audio-visual; fMRI, functional magnetic resonance imaging; BOLD, blood oxygen level-dependent; RT, reaction time; V, visual; 3D, three dimensional; Voluntary, voluntary eye movement task; Involuntary, involuntary visual task; Search, visual search task; ROI, region of interest; MNI, Montreal Neurological Institution; EPI, echo planner imaging; GLM, General Linear Model; ACT, activation; SD, standard deviation; SEM, standard error of the mean; IPL, inferior parietal lobe; IPS, intraparietal sulcus; SPL, superior parietal lobe; STS, superior temporal sulcus; TMS, transcranial magnetic stimulation; ITG, inferior temporal gyrus; SG, supramarginal gyrus; IFC, inferior frontal cortex.

* Corresponding author at: Eleanor Rathbone Building, University of Liverpool, Bedford Street South, Liverpool L69 7ZA, United Kingdom.
E-mail address: k.alwashmi@liverpool.ac.uk (K. Alwashmi).

https://doi.org/10.1016/j.neuroimage.2023.120483
Received 19 June 2023; Received in revised form 18 November 2023; Accepted 1 December 2023

1. Introduction and background

1.1. Systematic audio-visual training

In natural environments, objects and events typically generate a correlated input to multiple sensory modalities at once. Simultaneous multisensory signals are perceptually combined and can enhance cognitive processing. Visual search, for instance, was found to be speeded by adding a synchronous tone in healthy subjects (Fleming et al., 2020; Van der Burg et al., 2008) and stroke survivors with visual field loss (Keller and Lefin-Rank, 2010). Several studies on visual field defects also confirmed the development of compensatory strategies after multisensory stimulation in hemianopia patients (Grasso et al., 2016; Passamonti et al., 2009; Alwashmi et al., 2022).

One of the most effective techniques in rehabilitating visual field defects is visual scanning training to compensate for visual field loss. This is a purely visual paradigm, however, the use of human multisensory capabilities is a promising approach to recovering from unimodal sensory impairments (Alwashmi et al., 2022). Systematic AV stimulation, for example, induced a long-lasting amelioration of visual field disorders (Bolognini et al., 2005). Several imaging studies in humans demonstrated that AV multisensory integration process involves both superior colliculus and posterior cortical areas, including temporoparietal and posterior parietal areas (Dundon et al., 2015). Since most patients who have damage to the visual cortex still have intact multisensory brain areas, it might be possible to train retinotectal functions by AV stimulation during both chronic and acute stages of stroke (Keller and Lefin-Rank, 2010; Grasso et al., 2016; Dundon et al., 2015).

Multisensory AV training is more effective for visual learning than uni-sensory visual training: enhanced performance was reported for visual-only perceptual tasks when, during training, auditory signals were presented together with visual target signal - for motion detection (Seitz et al., 2006), for visual motion coherence detection (Kim et al., 2008), and visual motion discrimination. Visual discrimination improved more strongly for motion directions that were paired with congruent sound changes during training (Seitz et al., 2006; Beer and Watanabe, 2009; Wuerger et al., 2003).

There is compelling evidence for a complex interdependency between spatial location and temporal structure in determining the ultimate behavioural and perceptual outcome associated with paired multisensory (i.e., AV) stimuli (Grasso et al., 2016; Passamonti et al., 2009; Alwashmi et al., 2022; Stevenson et al., 2012; Evans and Treisman, 2010; Meyer and Wuerger, 2001). As a general rule, the more temporally synchronous the paired stimuli are, the greater the multisensory enhancement (Stevenson et al., 2012; Evans and Treisman, 2010; Frassinetti et al., 2002).

There are two factors that could influence multisensory processing and behavioural enhancement: multisensory perceptual learning and cross-modal attention.

1.2. Audio-visual perceptual learning and multisensory integration

The brain's perceptual ability continually evolves and changes throughout an individual's lifetime. These modifications can either be temporary or result from long-lasting perceptual learning, which can take place unconsciously (Sasaki et al., 2010; Lu et al., 2011).

Our perception of the world and our interactions with it are profoundly influenced by how we process sensory information. While our senses transduce information independently, they often work together to create a unified perception (Lauzon et al., 2022). Multisensory integration offers various advantages in terms of quicker and more accurate perception leading to improved behavioural responses (Stein and Meredith, 1993). Interestingly, the way in which sensory inputs from different modalities are integrated can be dynamically altered through perceptual training (Altieri et al., 2015). For instance, training involving perceptual feedback can narrow the temporal binding window, enhancing individuals' temporal precision when making judgments about audiovisual relationships (Powers et al., 2012; Stevenson et al., 2013; Powers et al., 2009). In addition, the auditory-visual spatial correspondence in crossmodal interactions may be altered by training or experience (Beer et al., 2011). In a recent investigation conducted by Lauzon et al. (2022), a direct connection was established between associative learning and the ability to merge information from multiple sensory modalities. Their findings revealed that individuals who demonstrated more robust indicators of associative learning also displayed more robust indicators of multisensory integration for the stimuli they had learned to associate (Lauzon et al., 2022). These examples underscore the adaptability of multisensory processing even in adulthood, demonstrating the brain's plasticity in merging information from different sensory channels.

Several studies provide evidence that multisensory integration occurs at low-level stages of sensory cortical processing and in subcortical structures (Beer and Watanabe, 2009; Tyll et al., 2011). Three possible mechanisms were suggested: feedback influences from multisensory convergence zones within higher association cortices or at the border of sensory-specific cortices, direct cortico-cortical interconnections between modality-specific areas, and integration of sensory information at subcortical levels, including sensory-specific thalamic nuclei (Tyll et al., 2011; Ghazanfar and Schroeder, 2006; Driver and Noesselt, 2008).

Common methods used to trace these mechanisms include single neuron recording, retrograde or anterograde tracing, functional magnetic resonance imaging (fMRI) studies, and electroencephalogram (EEG) or event-related potential (ERP) studies of certain regions (Murray et al., 2016). fMRI can be used in humans as well as with non-human animals and can provide information *in vivo* on a large population of neurons, which gives insight into high-level neural organisation (Murray et al., 2016).

Bonath et al. (2013) orthogonally manipulated temporal and spatial congruency of AV stimuli, subjects were tasked with judging either their temporal or spatial congruency. Using voxel-based fMRI, increased signals were observed in the posterior and central thalamus for the temporal and spatial tasks, respectively. These results indicate that both the AV stimulus configurations and task-related processing of temporal or spatial features can selectively modulate thalamic processing and influence cortical processing at an early stage (Bonath et al., 2013). Noesselt et al. (2010) looked at the effect of co-occurring sounds on behavioural visual detection sensitivity and neural responses to visual stimuli of varying intensities. The study's findings suggest that the multisensory enhancement of detection sensitivity is mediated by a brain network that involves not only the established multisensory areas like the superior temporal sulcus, and sensory-specific cortex like the primary visual or auditory cortex but also the visual and auditory thalamus.

Recent studies suggest that as the number of sensory inputs carrying significant information increases, the processing demands can be shifted to other cortical brain regions besides the sensory-specific cortex due to the increased complexity of such stimuli. Stickel et al. (2019), Porada et al. (2021), for example showed that activation in the left inferior frontal gyrus and left inferior parietal cortex was increased during bimodal stimulation and even more so during trimodal stimulation (olfactory, visual, and auditory).

Research has also shown the cerebellum's role in multisensory perceptual learning, indicating that its functional activation is linked to the perceptual requirements of a task. Cerebellar activation was found to be significantly correlated with increases in the perceptual demands of the task for both auditory and visual stimuli (Baumann and Mattingley, 2010), and a combined AV motion detection led to an increase in cerebellar activation compared to unimodal visual and auditory motion tasks (Baumann et al., 2015).

1.3. Cross-modal attention and multisensory integration

There are many features we can direct our attention to. One feature that is shared by both auditory and visual events is their location in space. Attentional processing can occur in a bottom–up (exogenous) manner, for instance, when a salient event pops out from its background (Van der Burg et al., 2008). In this case, an object is selected even though the observer was not planning to select it. In other cases, attentional processing operates in a top–down (endogenous) manner in which the observer voluntarily controls what is attended and what is not (Koelewijn et al., 2010). Simultaneous and co-localised AV cues enhance saliency and attract attention when individual stimuli are less effective. Both spatial attention and multisensory integration can take place in higher hetero-modal brain areas (e.g., superior colliculus, thalamus, superior temporal sulcus, and intraparietal areas) but also in early primary sensory areas (e.g., primary visual and auditory cortices) in parallel fashion (Koelewijn et al., 2010; Macaluso and Driver, 2005; Wuerger et al., 2012).

As attention can influence multisensory integration, it is not a purely automatic process (Koelewijn et al., 2010). Frassinetti et al. (2002) show that in multimodal brain areas, there may be an overlap between the physiological mechanisms responsible for stimulus-driven cross-modal spatial integration and mechanisms controlling stimulus-driven (exogenous) cross-modal spatial attention.

1.4. VR and audio-visual perceptual learning

The typical presentation of AV stimuli has been implemented by using separate speakers mounted on different spatial locations, which aligns the acoustic stimuli with the visual stimulus locations (Stevenson et al., 2012; Sürig et al., 2018; Van Wanrooij et al., 2009). The interaction between the spatial location and temporal structure of paired AV stimuli could be modulated by the complexity and ecological validity of the stimuli, which may reflect the statistical, semantic and contextual realities of real-world events (Stevenson et al., 2012).

Incorporating 3D audio is a crucial component of virtual reality (VR) applications. By simulating the position and distance of audio sources, 3D audio allows for more immersive and realistic experiences for users (MetaQuest, 2022). Audio can be played through headphones to indicate the location of the source in the virtual environment and direct the player's attention (MetaQuest, 2022). To create a truly realistic 3D audio experience, it is necessary to track the position of the audio source and simulate its surrounding environment. One 3D audio engine, Oculus Spatializer, uses a generic head-related transfer function (HRTF) to provide auditory spatialization (MetaQuest, 2022; Beig et al., 2019). Meyer et al. (2005) demonstrated that AV perceptual integration requires very high-quality localisation cues, and for facilitation of motion detection to occur, it is not sufficient to have local visual signals, but that a high quality auditory spatial signal is also crucial. However, research suggests that judging the absolute position of a sound source in 3D space is most accurate for horizontal location (azimuth), but less accurate for vertical location (elevation) (Beig et al., 2019; Danevičius et al., 2021).

Spatial and temporal coincidence are the dominant cues for integration, but the brain also relies on a feature correspondence between the different sensory inputs (Evans and Treisman, 2010). Pitch, for instance, is involved in several cross-modal correspondences. Bernstein and Edelstein (1971) were the first to measure the cross-modal congruence between pitch and visual vertical position, subsequently replicated by many other studies (Evans and Treisman, 2010; Bernstein and Edelstein, 1971; McCormick et al., 2018; Chiou and Rich, 2012). Participants classified visual stimuli as high and low more quickly when the visual stimulus was accompanied by a tone that was congruent rather than incongruent (e.g., high pitch with high position rather than low).

The brain has an ability to adapt to changes in sensory input to determine a sound's position by learning and calibrating these cues, using accurate spatial feedback from other sensorimotor systems (Hofman et al., 1998). Hofman et al. (1998) showed that the adult human auditory system undergoes ongoing spatial calibration when the spectral elevation cues of human subjects were disrupted by modifying their outer ears. This study found that although localization of sound elevation was significantly reduced immediately after the modification, accurate performance was gradually regained, suggesting that learning the new spectral cues did not interfere with the neural representation of the original cues, as subjects could localize sounds with both normal and modified pinnae. Therefore, learning to use novel auditory cues to identify visual target's location is of our interest.

VR is widely used in learning and rehabilitation applications (Seymour et al., 2002; Cooper et al., 2021; Cooper et al., 2018; Larsen et al., 2009; Huygelier et al., 2022). Cross-modal learning effects in a virtual environment can help determine how information is integrated across sensory modalities. The influence of sound on visual motion perception in a virtual reality scenario (Allue et al., 2016), AV cross-modality in the context of material appearance perception (Malpica et al., 2020), and the role of active multisensory-motor interactions when hearing individuals adapt to altered binaural cues (Valzolgher et al., 2020) have been investigated in previous literature.

To summarise, auditory cues aid visual learning. For optimal effect they need to be co-incident and co-localised with the visual target location (or motion). In addition, auditory stimuli can exploit natural feature correspondences, such as the association between sound pitch and perceived visual elevation. Participants can be expected to learn novel signal characteristics, for example the generic HRTFs used in VR signalling and the pitch-visual correspondence, over time. This learning of AV association is expected to aid faster response time in a trained AV voluntary eye movement task with specific eye movements strategy, that may also apply to untrained tasks when integrating the same auditory cues with different eye movement mechanisms.

1.5. Study aims

How multisensory perception is altered in an immersive virtual environment is an area that has received some interest, (Allue et al., 2016; Malpica et al., 2020; Valzolgher et al., 2020) but many questions remain. This study aimed to investigate the effects of a multisensory training program utilising VR on brain activity and cognitive performance. Here we examined the effect of systematic AV training by combining horizontal auditory spatialisation and vertical pitch-position in congruent with visual targets. We applied pre/post-training comparisons between bimodal and unimodal tests. In addition, we investigated the learning transfer of the voluntary systematic AV stimulation to an involuntary visual task and a visual search task. A better understanding of neural mechanisms underlying multisensory integration in virtual environments and its contributions to learning can play a key role when implementing training protocols for rehabilitation and for the development of assistive technologies.

2. Materials and methods

2.1. Participants

Twenty right-handed healthy participants (7 female,13 male) with normal or corrected to normal vision and normal hearing were recruited (mean age = 25.4 years, range 20–37). All participants gave informed consent, and the study was approved by the University of Liverpool Ethics Committee (reference number:6446).

2.2. VR training-game design

All participants followed a VR implementation of an eye movement training programme developed for hemianopia neurorehabilitation: VISION visual scanning training (Aloufi et al., 2021; Rowe et al., 2017).

Hemianopia is a complete or partial blindness in the visual fields of both eyes, commonly caused by cerebral infarction (Alwashmi et al., 2022). Visual scanning training is designed to help participants to scan their environment in a systematic fashion. To this end targets in a predictable sequence of positions alternate between the left and right visual hemifields. The basic visual training concept remained unchanged, except that audio cues were added, and the systematic AV training was incorporated into a virtual reality game.

The game was created using Unity 2018.0.3f and programmed in C#, designed for the Oculus Quest 2 (Meta Quest 2, 2022). The game was designed for the player to be in a seated position for participant safety and comfort. Participants were required to use the VR headset for training in a quiet space away from other people or objects. To obtain behavioural performance measures, participants were instructed to respond using two different buttons to two different types of targets, identified by geometric symbols. The game consists of two levels of adaptive difficulty as increasing the challenge has been shown to motivate the user (Cavaco et al., 2015). The goal of the game is to maximise the number of points obtained in a 30 min interval by maintaining high accuracy and faster response time.

2.2.1. Visual stimuli

During the game, different shapes are spawned and move toward the player. Two targets (triangle and circle) were embedded in 5 types of distractors (square, diamond, cylinder, vertical rectangle, and a horizontal rectangle). Two or more shapes appear simultaneously, but on opposing angles on x, y planes, one is the target, and the rest are distractors.

The systematic spawning of targets is done by along three axes on the y plane, elevation ranges between +- 14°, azimuth +- 45, +- 90° for levels 1 - 2, respectively, for visual field expansion. The y coordinates for the first axis starts from the centre of player azimuth with a small offset and have the same value, the remaining two axes are diagonal to the first and are linearly spaced apart, starting from the top left - bottom right and top right - bottom left of a curved rectangle, and are instantiated in a semi-circle around the player. A picture (Fig. 1) demonstrating how visual targets appear is adapted from Aloufi et al. (2021). The targets radiate from the inside out from 42 positions and gradually move towards the player at their fixed angle (Video-clip in supplementary material).

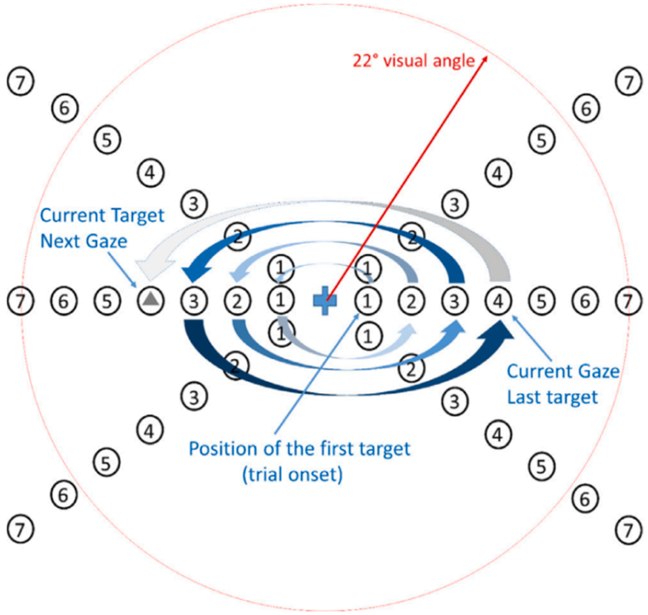

Fig. 1. Schematic diagram of the stimulus sequence used in training: Participants learn to rapidly gaze at alternating target positions, starting with 1, gradually moving more eccentric to positions marked 7. At each position, they are asked to identify a target symbol (triangle or circle). The targets are small so that participants have to foveate. Horizontal and diagonal movements are practised in sequence. From Aloufi et al. (2021).

The player is instructed to press two different buttons (A for circle, or B for triangle) on the Oculus controller. The game simulates a space scenario where participants are faced with a random sequence of targets. Circles and triangles bring rewards when identified correctly and penalties when errors are made. Pressing the 'A' button raises a red shield for the circle, while pressing the 'B' button raises a green shield for the triangle which pull in the target. A new set of distractors is displayed for every new trial.

Players gain or lose 10 points depending on their responses, and the total score is displayed during the game to maintain motivation. The participants are placed into level two automatically after ten completed sessions. The game terminates after 30 min for each session.

2.2.2. Audio cues

The AV stimuli were spatially and temporally congruent, so that the audio cues were played when a target is spawned and lasted 100 ms (Keller and Lefin-Rank, 2010; Bolognini et al., 2005). We used procedurally generated noisy sinusoidal tones, as Unity's sound engine was not capable of generating audio cues that could be reliably localised (Cavaco et al., 2015). The Oculus Sound Spatializer (MetaQuest, 2022) was used to localise tones on the azimuth, which simulates appropriate interaural time and level differences, while the vertical target position was encoded using frequency (3000 Hz–9000 Hz). This range was chosen as the results from Risoud et al. (2018) indicate that sounds above 3000 Hz provide better localisation. Frequency on the vertical plane is also associated with higher-pitched tones (Evans and Treisman, 2010; McCormick et al., 2018; Chiou and Rich, 2012; Risoud et al., 2018; Roffler and Butler, 1968; Pratt, 1930; Mudd, 1963). Auditory tones were programmatically generated for each visual target position. To further aid in localisation, the generated signals were created with 20 % white noise ratio. The initial signals S were generated with the following formula:

$$S_i = v(2N_i - 1) + \left(a * \sin\left(\frac{2\pi i f}{r_s}\right)\right)$$

Equation 1. equation representing the noisy sinusoidal waveform.

Where i is the current time index for the signal S, v is the amplitude (0.20 here for a 14 dB SNR) of a uniformly distributed noise, a is the amplitude of the sinewave signal (1 here), f is the desired frequency for the signal, and rs is the desired sample rate (44,100 Hz here). The Oculus Sound Spatializer then modified the signal to add positional information that used head-related transfer functions (HRTFs), which added interaural time delay and interaural level differences into the generated signals, consequently adding in the directional information.

2.2.3. Training paradigm

Participants were asked to train at home for 30 min daily, five days a week and for four consecutive weeks (two weeks on each level) (Fig. 2). The decision to require participants to engage in 30 min of daily training at home for four weeks is based on established practices in the field, supported by prior literature (Aloufi et al., 2021; Rowe et al., 2017). Previous studies with similar training protocols have shown that this timeframe is effective. For example, a study on visual search training for individuals with hemianopia found significant improvements in search efficiency after four weeks of training (Mannan et al., 2010). Additionally, our choice of a four-week duration aligns with research showing early functional and structural changes after just two weeks of training, which subsequently correlate with behavioural improvements after six weeks (Aloufi et al., 2021). Performance data (response times, accuracy, and total score) were recorded automatically and checked after each level (every two weeks).

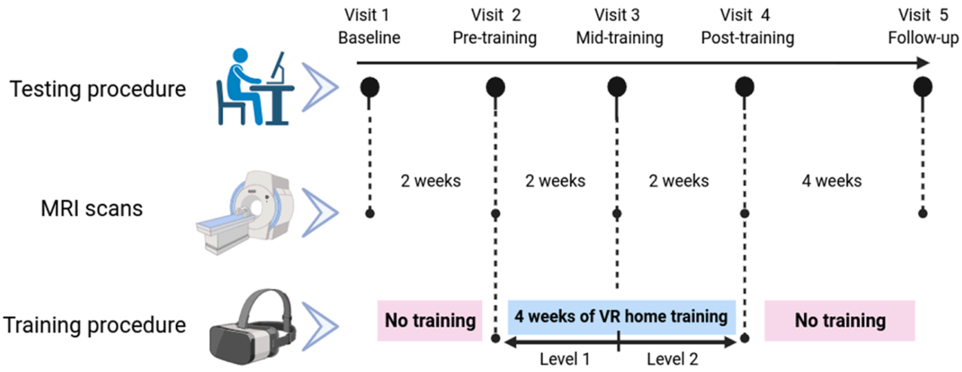

Fig. 2. Data collection methodology: this chart shows the timeline of our study for the training procedure at home on the VR headset, the testing procedure in the lab, and the MRI scanning over five visits. (Figure is created in BioRender.com).

2.3. *Testing paradigm*

In addition to the performance monitoring that was conducted while participants were training at home, each participant was assessed in the laboratory every two weeks. A digital version of the VISION visual scanning training card (Aloufi et al., 2021) was implemented in PsychoPy2 *(v1.82.01;* (Peirce et al., 2019)) to record the response time and accuracy. Recorded audio cues drawn from the VR game, and matched the spatial location of the visual targets, were applied to the test. An eye tracker (The EyeTribe – Model: ET1000), was used to monitor the visual performance. The data gathered at home were used to check compliance with the training regime and improvement over time on the VR game. The data recorded in the controlled laboratory environment were used to quantify learning gain and its effect on the other tasks (see testing procedure Section 2.3.1).

2.3.1. *Testing procedure*

At the first visit each subject performed a familiarisation block, through which participants were allowed to ask questions and practice all the tasks before their first MRI scan. Therefore, the first one third of the data was removed from the first behavioural test and the remaining trials was considered as a baseline. Behavioural performance and MRI scans were measured consecutively; pre-training, after 2 weeks of training, post-training (after 4 weeks of training) and at a follow-up test, one month after the training ended, to quantify AV learning gains (Fig. 2).

All participants performed a session of the same 252 targets at each data collection point. The test session was performed on a laptop screen of 38.4 × 21.6 cm (width x height) at a viewing distance of 50 cm. A display mounted eye tracker individually calibrated for each test session, was used to record gaze position data. JBL Tune 750BTNC Noise-Cancelling Headphones were used for AV tests.

For every laboratory session, all participants underwent three tests. Each test was conducted both with audio cues (AV) and without audio cues (visual (V)) to compare the two conditions. Multisensory integration can interact with attention. Tang et al. (2016) explained that attention could modulate multisensory processing in two ways; the voluntary (endogenous or top-down) is associated with the presentation of cues that indicate where, when, or to what subjects should direct their attention (Tang et al., 2016). The involuntary (exogenous or bottom-up) is associated with the presentation of salient targets, particularly when they appear in unexpected locations.

Here, we compared tasks that could represent both endogenous and exogenous ways. The first test employed was previously described as a voluntary systematic eye movement task (voluntary) (Fig. 1) (Aloufi et al., 2021), through which the subjects can direct their attention to targets that appeared in a predictable sequence of locations in the visual field. The second was the exogenous task (involuntary), where a salient targets either circle or triangle (gray colour shapes with random size variation between 0.92° and 1.14° visual angle) presented on a blank background in random positions. The third was the visual search task (search) where the same card used in the voluntary task (Fig. 1) (Aloufi et al., 2021) was employed, but targets appeared randomly, in one of the 42 locations on the card, rather than in a predictable sequence.

2.3.2. *Behavioural performance measurement*

The principal behavioural performance measure was the mean response time (RT): the average time taken by participants to respond to each stimulus in the experiment. For experiment conducted in the lab, the number of targets in each run was fixed, so that the time spent on the task varied each visit. For the training on the VR runs at home, the total execution time was fixed (30 min) to ensure fixed training period during the daily session, so that the number of targets processed varied between participants. The fMRI experiment was a block design, with alternating ‘task’ and ‘rest’ conditions, here the total execution time (task block duration) was fixed so that, as in the VR, the number of targets processed varied between participants and MRI sessions. Learning-related behavioural changes in this study were expected for response time rather than for response accuracy, where we expect near ceiling performance.

2.4. *Stimulus delivery in the scanner*

Participants were positioned inside the MRI scanner in a dimly lit room and instructed to remain still. The stimuli were displayed on an MRI compatible monitor (NordicNeuroLab, Model LCD 3.0.4, (NordicNeuroLab, 2023)). Participants viewed the stimuli through a mirror attached to the head coil, providing a visual field of 45°. PsychoPy2 (v1.82.01) (Peirce et al., 2019) was utilised for stimulus presentation and to document the participants' responses through the NordicNeuroLab fMRI Response Grips (with the right index finger being used for circle targets and right thumb for triangle targets).

2.5. *Functional imaging data analysis*

Our imaging analysis was divided into two parts. First, we conducted a whole-brain analysis to identify overall task-relevant activation patterns and significant changes in functional brain activity following a four-week training program. This analysis included all brain regions to detect any substantial alteration in multisensory and sensory-specific cortices. While hypothesis-driven voxel-wise ROI analysis can predict changes in training-induced activity and reduce exploratory analyses in a whole-brain analysis (Erickson et al., 2007), it is limited to specific voxels, and may not capture training-induced changes that occur in other regions of the cortex. Furthermore, analysing pre-defined voxels instead of a larger ROI may not accurately reflect changes throughout the rest of that region (Erickson et al., 2007).

After this analysis, we examined the time-course of changes in the functional beta values (β) and focused on task-relevant regions identified in the previous step. The final analysis aimed to determine if there was a correlation between neuroimaging measures and behavioural performance, and if it was feasible to predict measures over time.

2.5.1. Functional imaging tasks

The presentation of stimuli and instructions for the participants were identical to those used during the training phase. To demonstrate the brain functional activation changes associated with learning, three tasks were presented during the fMRI scans. These tasks included the AV voluntary task, representing the trained task, the V voluntary task for comparing visual and audiovisual conditions, and the AV involuntary task to compare between AV voluntary (endogenous) and AV involuntary (exogenous) tasks (refer to Section 2.3.1 Testing Procedure for task details). Global activation patterns during the execution of the AV voluntary (trained) task were compared with those observed during the other two tasks.

Each task was structured as 15 blocks with a block duration of 32 s, alternating 16 s of rest and 16 s of activity, in line with the block design described in Maus et al. (2010). The completion time for each functional task was 8 min (320 vol). Participants were asked to focus on a central fixation point during the rest periods. The accuracy of responses and the total number of responses to targets were recorded using PsychoPy2 (Peirce et al., 2019).

2.5.2. fMRI data acquisition

In this study, both T1-weighted anatomical images and T2*-weighted functional MR images were obtained using a 3T Siemens Magnetom Prisma and a 32-channel receiver head-coil. The T1-weighted images (mprage) had a voxel size of 1.0 × 1.0 × 1.0 mm, a repetition time of 2000.0 ms, an echo time of 2.25 ms, and an inversion time of 912 ms. The functional imaging data was captured using a multi-slice echo-planar (EPI) pulse sequence (a T2*-weighted gradient-echo). The EPI iPAT (integrated Parallel Acquisition Technique) acceleration factor was set to 2, and the repetition time was set to 1500 ms with an echo time of 30 ms and a flip angle of 90∘. The acquisition covered the entire brain volume using 48 interleaved transverse slices with a voxel size of 3.0 × 3.0 × 2.7 mm and 10 % distance factor. The field of view was 192 × 192 mm.

2.5.3. fMRI statistical analysis

The functional data analysis was carried out using SPM12 (SPM-12, 2023), a Statistical Parametric Mapping software that operates in MATLAB R2018b. The pre-processing of the functional data was completed using the default batch in SPM12 (preproc_fMRI.m) with some changes to meet the nature of multi-slice sequence. The procedure started with slice time correction and then realigned all the functional scans to a common image. For the slice-timing correction step, we put-in the slice-timing vector instead of slices order for all volumes and we enter the reference time in milliseconds for the reference slice. For each participant, the pre-processing was performed on all fMRI scans of the AV voluntary, V voluntary, and AV involuntary tasks that were conducted at each of the five visits (three runs per visit). During the pre-processing, each image within a session was spatially realigned to the first volume of the session, corrected for motion artifacts, and unwrapped. The T1-weighted (mprage) structural image, which was obtained during the baseline scan, was segmented into white matter, gray matter, and cerebrospinal fluid to estimate spatial normalisation parameters from the subject's native space to Montreal Neurological Institution (MNI) space. All EPI images for each participant were then coregistered to this high-resolution T1 structural scan. To conclude the pre-processing, all the coregistered EPI images were normalised to the MNI space using the estimated normalisation parameters, resampled to 1 × 1 × 1 mm3 voxels using trilinear interpolation, and spatially smoothed with a 6 mm full-width at half-maximum Gaussian kernel for group analysis.

2.5.4. fMRI data analysis

The analysis of the fMRI data was performed using a two-level procedure with a random-effects model (Friston et al., 1999). The first level analysis involved modelling the fMRI signals for each participant using a General Linear Model (GLM) and a design matrix. The matrix included the onsets and durations of each block in the AV and V voluntary, and AV involuntary. Six regressors were used, including rest_voluntary (AV), act_voluntary (AV), rest_voluntary (V), act_voluntary (V), rest_involuntary (AV), and act_involuntary (AV). From these regressors, three contrasts were created for each visit (act_voluntry(AV) vs rest_voluntary (AV), act_voluntary(V) vs rest_voluntary(V) and act_involuntary vs rest_involuntary), to determine the task-relevant brain regions in the fMRI data for further analysis.

The second level, a group analysis, involved comparing the baseline and post-training BOLD signal activation maps using paired T-tests. This analysis was performed separately for each of the three functional tasks (voluntary_AV, voluntary_V, and involuntary_AV). The analysis consisted of two contrasts for each task to compare activation post-training (visit 4) with baseline; ACTv1 > ACTv4 and ACTv1 < ACTv4. The results were thresholded at pFWE < 0.05, with cluster-level probability values computed from a voxel-level threshold of p-unc < 0.001.

2.5.5. fMRI data extraction for time course analysis

The analysis of learning induced changes over time was based on brain regions that displayed significant functional changes at the post-training visit compared to the baseline in the trained bimodal AV voluntary eye movement task. The MarsBar ROI toolbox for SPM (release 0.42) (Brett et al., 2002) and REX (2023) were utilised to extract the mean activation in these regions across all other time points. The anatomical locations of these areas were determined using the SPM-Anatomy toolbox v.1.7 (atlas-based). The regions included the thalamus, cerebellar crus I regions, and the inferior parietal lobe (IPL). In case multiple activation clusters were present within a single anatomical region, they were combined.

Repeated measures ANOVA was performed using OriginPro for windows (2020b Version 9.7.5.184) to evaluate brain activation change over time. Sphericity was tested using the Mauchly's Test, and results were corrected using Greenhouse-Geisser correction when probability value of approximate ChiSq was less than 0.05 indicating that the assumption of sphericity had been violated (Singh et al., 2013). The correlation between changes in behavioural performance and functional activation patterns was also investigated. Following previous literature on motor and perceptual learning tasks (Yotsumoto et al., 2008; Furmanski et al., 2004; Kourtzi et al., 2005; Tombari et al., 2004) two-sided *t*-tests were performed to test for functional activity changes with training.

3. Results

3.1. Behavioural data

All 20 participants completed the training period with the minimum daily training session (30 min daily) and exceeded the minimum performance requirement (70 % correct target identification in each task session) with mean accuracy in the VR home training 95.4 % (SD ±4.3) and in the lab tests 94.2 % (SD ± 6.9). Thus, all behavioural data were included in the data analysis. Normality checks were carried out on the dependent variable by task, which were approximately normally distributed. Separate repeated measures ANOVAs with a Greenhouse-Geisser correction (Greenhouse and Geisser, 1959) was applied whenever indicated by Mauchly's test of sphericity; corrected *p*-values and degrees of freedom are reported. Bonferroni correction was used for multiple *post hoc* tests.

3.1.1. Impact of training – VR behavioural data

Behavioural performance improved significantly over the training period. In the RT group analysis, there is a significant main effect of session, number of distractors, and field of view between levels of difficulty in a repeated measure ANOVA (session: $F_{(19,361)} = 11.454$; $p < 0.0001$, number of distractors: $F_{(19,361)} = 6.84$; $p < 0.0001$, field of view: $F_{(19,361)} = 11.82$; $p < 0.0001$, respectively). *Post hoc* comparisons revealed that the mean RT decreased significantly from 1.63 (±0.4) s at day one to 1.35 (±0.3) at day 10, and further decreased to 1.23 (± 0.14) s at day 20 ($t_{(19)} = 6.7$; $p < 0.0001$, $t_{(19)} = 9.24$; $p < 0.0001$, respectively) (SupFig. 1.a). Significant increase in the mean total score was found from 5819.812 (±939.390) at day one to 7633.125 (±869.627) at day 20 ($t_{(19)} = 11.82$; $p < 0.0001$) (SupFig. 1.b), with a significant main effect of session in a repeated measure ANOVA ($F_{(19,361)} = 16.29$; $p < 0.0001$).

Thurstone (1919), approximately a century ago, made an insightful observation that learning curves tend to exhibit erratic behaviour, necessitating the examination of overall trends across multiple observations rather than individual data points (Thurstone, 1919). Considering this, they proposed an exponential decay function as a suitable model for describing learning behaviour. Learning curves have proven to be valuable tools for quantifying and monitoring the performance of individuals engaged in novel or repetitive tasks (Anzanello and Fogliatto, 2011), as they capture learning behaviour and facilitate comparisons beyond isolated time points (Cooper et al., 2021). As the game visual angle was expanded from 45° to 90° moving from level 1 to level 2, we were interested in assessing whether the learning improvement was maintained in the first visual angle even when the angle been expanded in the second level. The mean RT in the 45° visual angle among the two levels was obtained and compared to the mean RT in the eccentric visual field in level 2. We employed an exponential decay function (Eq. (2)) to fit the performance data obtained from the VR over four weeks of training (20 sessions) on the first visual angle (45°) and over two weeks of training (10 sessions) on the second visual angle (90°) separately (Fig. 3).

$$y = A1 * exp - (x - x0)/t1 + y0$$

Equation 2: Where $x0$ and $y0$ are the x and y offset respectively, $A1$ is the initial amplitude and $t1$ the time constant defining the exponential decay.

For the first training level (all targets within 45° visual angle), the y offset ($y0$) was estimated to be 1.17 ± 0.03. The parameter $A1$, representing the initial amplitude (time to complete the task), was estimated as 0.39 ± 0.02. The time constant ($t1$) of the decay was estimated to be 9.79 ± 1.96. The goodness-of-fit measures indicate a high level of agreement between the fitted curve and the data, with a reduced chi-square value of 3.66e-4, an R-Square (Coefficient of Determination) of 0.96 for the first visual angle and a reduced chi-square value of 3.91e-4, an R-Square of 0.9 for the second visual angle. These results suggest that the exponential decay model effectively captures the observed decrease in mean RT over the 20 training sessions and that the learning continues for the centric visual targets even when the visual angle expanded in the last two weeks of training.

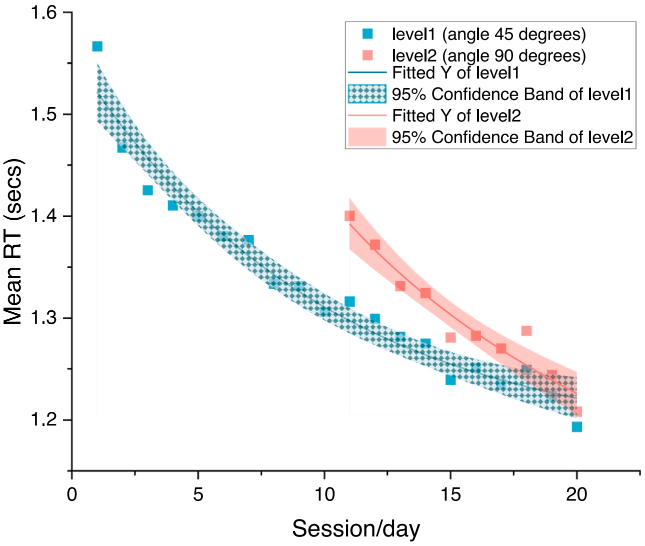

Fig. 3. Mean completion times across two levels of difficulty in all 20 sessions fitted with exponential curve: The bands represent 95 % confidence intervals. The colours indicate the two visual angles used in each level (light blue = level 1 & 2 with 45° visual angle, light red: level 2 with 90° visual angle). RT: response time.

3.1.2. *Impact of training – lab behavioural data*

The behavioural data were collected over five visits in the laboratory. Subjects trained on the VR system for four continuous weeks. The final data set was collected one month after the end of training as a follow-up session. Post hoc comparisons were conducted between baseline and post-training, both sessions were also compared to follow-up (visit 5) session. Response time outliers that defined as elements more than three standard deviations from the mean were removed before carrying out the statistical analyses (MATLAB rmoutliers, 2023). In our analysis, the primary focus was on assessing the impact of the training intervention. Consequently, the baseline and post-training assessments were considered as the main results of interest, which were presented in our figures.

Participants were tested twice before the training started, once during recruitment and once immediately before the first training session. Notably, when comparing mean response times between the baseline and pre-training assessments in the AV condition for all three tasks (voluntary, involuntary, and search), we found no significant changes (voluntary: $t_{(19)} = 2.19, p = 0.32$; involuntary: $t_{(19)} = 0.7, p = 1$; search: $t_{(19)} = 2.06$, $p = 0.42$). As both the baseline and pre-training assessments occurred prior to the actual training, we only report performance measures for the baseline test. Behavioural figure displaying all five visits are included in the supplementary materials (supFig. 2) to provide interested readers with the opportunity to examine the data across all time points.

3.1.2.1. *Voluntary systematic eye movement task (AV voluntary-trained and V voluntary-untrained).*

AV and V behavioural performance improved significantly over the training period. In the group analysis, A two-way repeated measure ANOVA was conducted for the voluntary task, utilizing time (visits) and condition (AV and V) as factors, with participants serving as the repeated measure. This analysis revealed significant main effect of time and the interaction between time and condition (AV vs. V) (Time: $F_{(2.35,\ 44.58)} = 92.03$, $p < 0.0001$; Interaction: $F_{(2.3,43.66)} = 3.11$, $p < 0.05$). In the *Post hoc* comparisons for each condition, the mean RT decreased significantly from 0.774(±0.022) s at the baseline to 0.556(±0.06) s post-training ($t_{(19)} = 11.95$; $p < 0.0001$) in the AV condition, compared to the non-auditory condition, where the mean RT decreased from 0.76 (±0.022) s at the baseline to 0.591 (±0.01) s post-training ($t_{(19)} = 10.42$; $p < 0.0001$) (Fig. 4).

To directly compare the audiovisual (AV) and visual-only (V) conditions in the voluntary eye movement task, a pairwise *t*-test was executed. The results of this test demonstrated a significant difference between the two conditions after two weeks of training (mid of training) as well as post training, with the response time in the audiovisual condition being notably faster than in the absence of audio cues (Mid of training: $t_{(19)} = -3.44$, $p < 0.01$, post training: $t_{(19)} = -6.86$, $p < 0.0001$) (Fig. 4).

This result shows that contrary to the baseline phase where no significant differences were observed between the AV and V conditions ($t_{(19)} = 0.263$, $p = 0.795$), there was indeed a notable difference in response times between these conditions in the voluntary task during and after the training period. This highlights the significant impact of audiovisual cues on enhancing task performance.

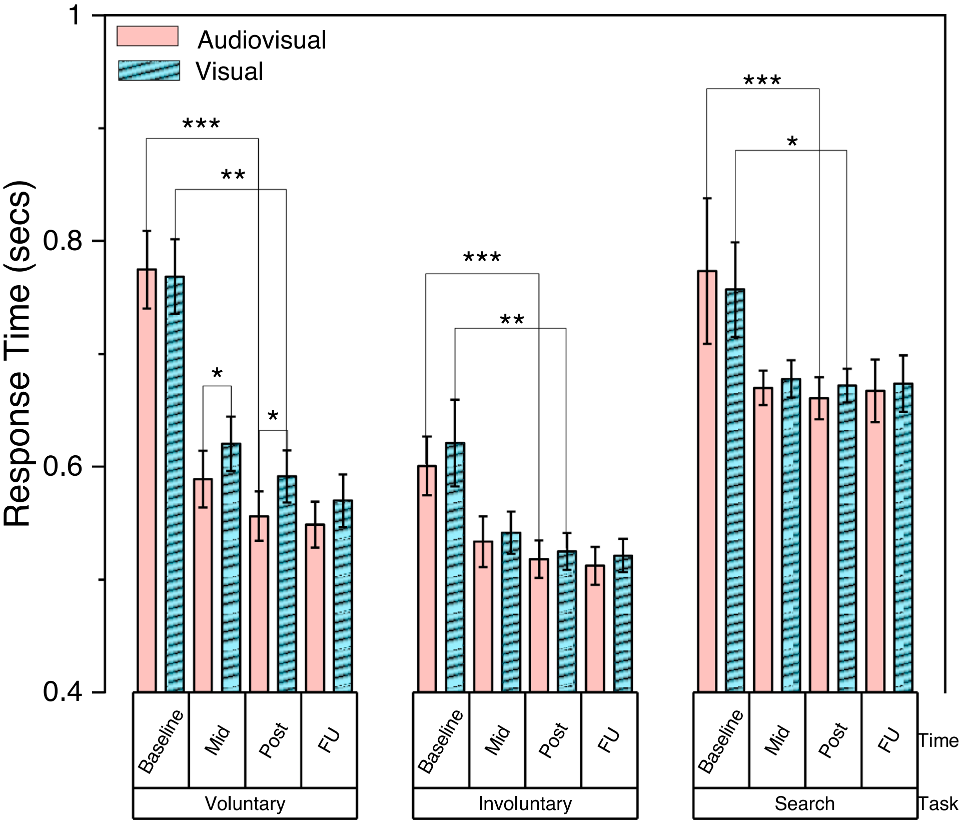

Fig. 4. Behavioural performance during the lab tests over time: this chart represents the mean response time (RT) change in three tasks (a-voluntary eye movement, b-involuntary eye movement, c-visual search) over four time points (Baseline, mid of training 'after two weeks of training', post-training 'after four weeks of training', follow-up (FU) one month after training ended). Light red: AV task, light blue: visual task without auditory cues. Error bars standard error of the mean (SEM).

3.1.2.2. Involuntary randomised task (involuntary- untrained task). The behavioural performance improved significantly over the training period. In the group analysis, a significant main effect of training duration was found in a two-way repeated measure (subject) ANOVA ($F_{(1.75,33.22)}$= 25.2; $p < 0.0001$), with a significant effect of interaction between time and condition (AV vs. V) ($F_{(2.3,\ 43.88)}$= 3.4; $p < 0.05$). However, no significant effect of condition was obtained ($F_{(1,19)}$= 1.3; $p = 0.26$). *Post hoc* comparisons for each condition revealed that the mean RT decreased significantly in the AV condition from 0.6 (±0.01)s at the baseline to 0.518(±0.01) s post-training ($t_{(19)} = 7.38$; $p < 0.0001$), compared to the non-auditory condition, where the mean RT decreased from 0.62(±0.03)s at the baseline to 0.525(±0.01)s post-training, where a significant change was, also, recorded ($t_{(19)} = 6.63$, $p < 0.0001$) (Fig. 4).

3.1.2.3. Visual search task (search- untrained task). The behavioural performance improved significantly over the training period. In the group analysis, there was a significant main effect of time in two-way repeated measure (subject) ANOVA ($F_{(1.2,22.78)}$= 6.89; $p < 0.05$). However, no significant effect of condition or interaction was obtained (condition: $F_{(1,19)} = 0.087$; $p = 0.77$, interaction: $F_{(1.81,\ 34.48)} = 1.86$; $p = 0.17$). In the *Post hoc* comparisons for each condition, the mean RT decreased significantly from 0.77(±0.04)s at the baseline to 0.66 (±0.01)s post-training ($t_{(19)} = 5.002$; $p < 0.001$) in the AV condition. Additionally, a significant change was recorded in the non-auditory task, where the mean RT decreased from 0.756 (±0.02)s at the baseline to 0.672(±0.01)s at post-training ($t_{(19)} = 3.78$; $p < 0.05$) (Fig. 4).

3.1.2.4. Follow-up behavioural measures. Behavioural performance improvements were maintained one month after training in all of the three tasks (voluntary eye movement, involuntary, and visual search). Mean RT in the follow-up session was significantly lower than baseline but not different from the RT measured post-training. The AV voluntary task (trained) and all the untrained tasks showed that the improvement was maintained in the presence as well as the absence of auditory cues. (Fig. 4, SupTable 1).

3.2. fMRI results

The purpose of this fMRI study was to examine the effect of auditory cueing of alternating speeded voluntary eye movements on brain activation. Participants underwent a baseline scan and a pre-training scan, with a 2, 3-week interval between the scans, before starting the training. As mentioned before, the primary focus was on assessing the impact of the training intervention. Consequently, the baseline and post-training assessments were considered as the main results of interest, which were presented in our figures. The hypothesis was that training would result in increased activation in task-relevant brain regions during the AV voluntary trained task. We, further, measured the effect of training on brain activation for two (untrained) tasks: AV involuntary and V voluntary eye movements (without audio cues).

The results showed that both the AV voluntary and AV involuntary tasks elicited similar global brain activation patterns (Fig. 5). However, the V voluntary task exhibited a clear difference in activation, with the temporal gyrus including primary auditory cortex not being activated.

After training, the AV voluntary (trained) task caused a significant increase in fMRI signal (Fig. 6). In contrast, no systematic training-induced increase in functional activity was observed in either of the untrained tasks. Instead, a reduction in signal activation was observed in different brain regions in these two tasks. No significant changes in signal activation were recorded between the baseline and pre-training scans, that performed before the training started. Therefore, for clarity

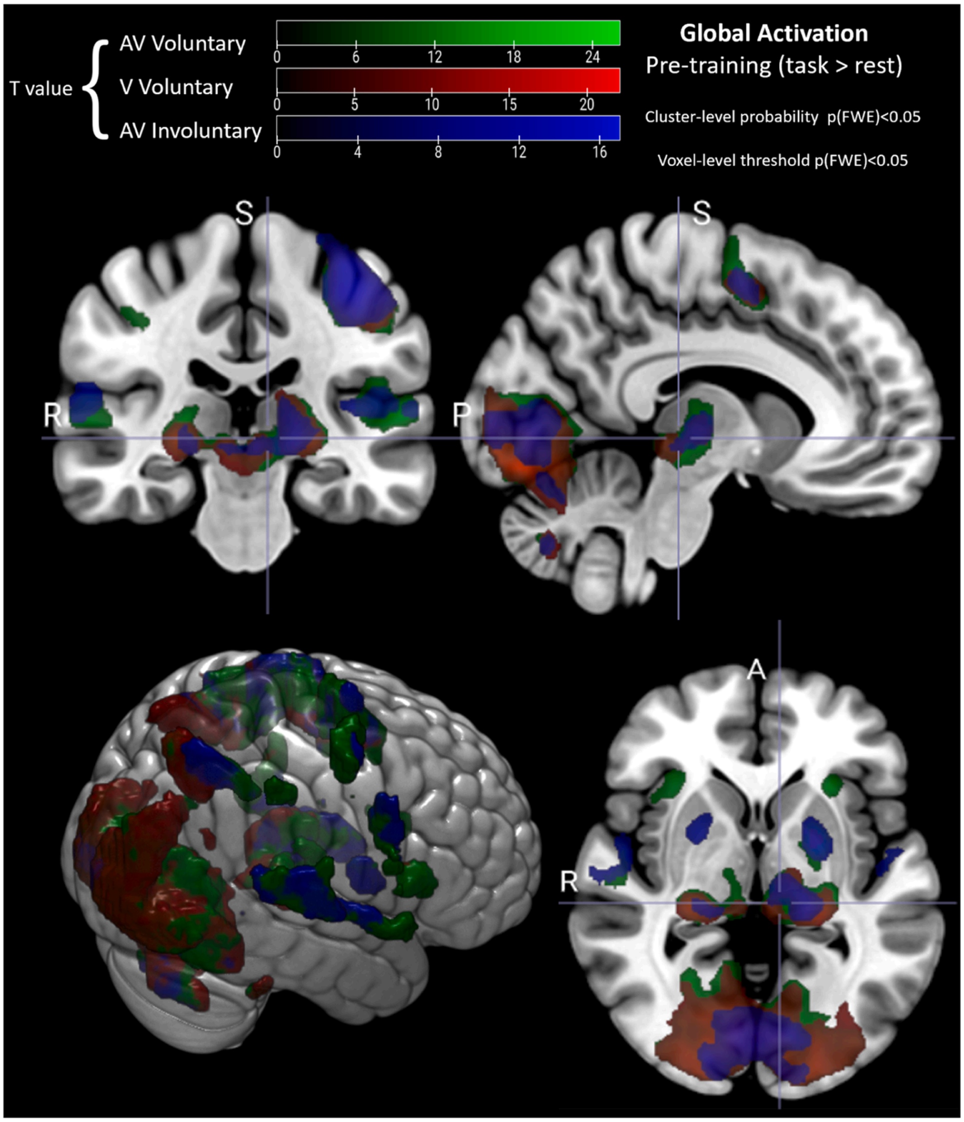

Fig. 5. Global functional activation for the trained (bimodal voluntary) and untrained (unimodal voluntary and bimodal involuntary) tasks: this figure shows the average global activation (p (FEW) < 0.05 cluster level) at baseline (pre-training) overlapped for the three tasks, where green represent the signal activation for the AV voluntary eye movement task, red: V voluntary eye movement task, and blue: AV involuntary eye movement task.

purposes, we only report fMRI measures for the baseline test (a figure including all time points can be found in the supplementary materials-supFig. 3).

The results of paired T-test (baseline and post-training) revealed a statistically significant increase in fMRI activation in the thalamus, the caudal inferior parietal lobe and cerebellum (Fig. 6, Table 1), with a *p* (FWE) of less than 0.05, while participants executed the AV voluntary trained task. Conversely, the inferior temporal gyrus (ITG) showed a decrease in fMRI signal during both untrained tasks, *p*(FWE) < 0.05 (Fig. 6, Table 1). Additionally, the AV involuntary task demonstrated a decrease in fMRI activation in the supramarginal gyrus (SG) and the inferior frontal gyrus (IFG), *p*(FWE) < 0.05 (Fig. 6, Table 1).

3.2.1. *Time course of the functional change (activation increase)*

The time course of functional activation changes for the three tasks was analysed using two-way repeated measures ANOVA, time (visits 1–5) and condition (AV-voluntary, V-voluntary, and AV-involuntary) as factors, and participant as a repeated measure, for each of the three regions where the training effect showed significant activation increase (thalamus, cerebellum, and IPL).

The results showed significant interactions between time and condition in the thalamus and cerebellum ($F_{(4.71,89.5)}$=2.82; $p < 0.05$, $F_{(8152)}$=2.02; $p < 0.05$, respectively). Meanwhile, cerebellum showed a significant main effect of time ($F_{(4,76)} = 3.54$; $p < 0.05$) and IPL showed a significant main effect of condition ($F_{(2,38)} = 13.3$; $p < 0.0001$) (supTable 2). Post-hoc, Bonferroni corrected, paired *t*-tests showed that the activation increases between baseline and post-training were significant in all three regions only for the trained AV voluntary eye movement task (Thalamus: $t_{(19)} = 4.01$; $p < 0.01$, cerebellum: $t_{(19)} = 5.2$; $p < 0.0001$, IPL: $t_{(19)} = 3.75$; $p < 0.001$) (supTable 2, Fig. 7). The activation during the AV voluntary task remained significantly different from baseline at the follow-up test one month after training was completed (Thalamus: $t_{(19)} = 3.06$; $p < 0.05$, cerebellum: $t_{(19)} = 3.5$; $p < 0.001$, IPL: $t_{(19)} = 3.14$; $p < 0.05$), which is consistent with the behavioural gains one month after training terminated.

3.2.2. *Time course of the functional change (activation decrease)*

The reduction in activation in untrained tasks was studied in three regions, inferior temporal gyrus (ITG), supramarginal gyrus (SG) and inferior frontal cortex (IFC). The study used the same two-way repeated measures ANOVA as described above. The results showed significant interactions between time and condition in the SG and IFC ($F_{(8152)} =$

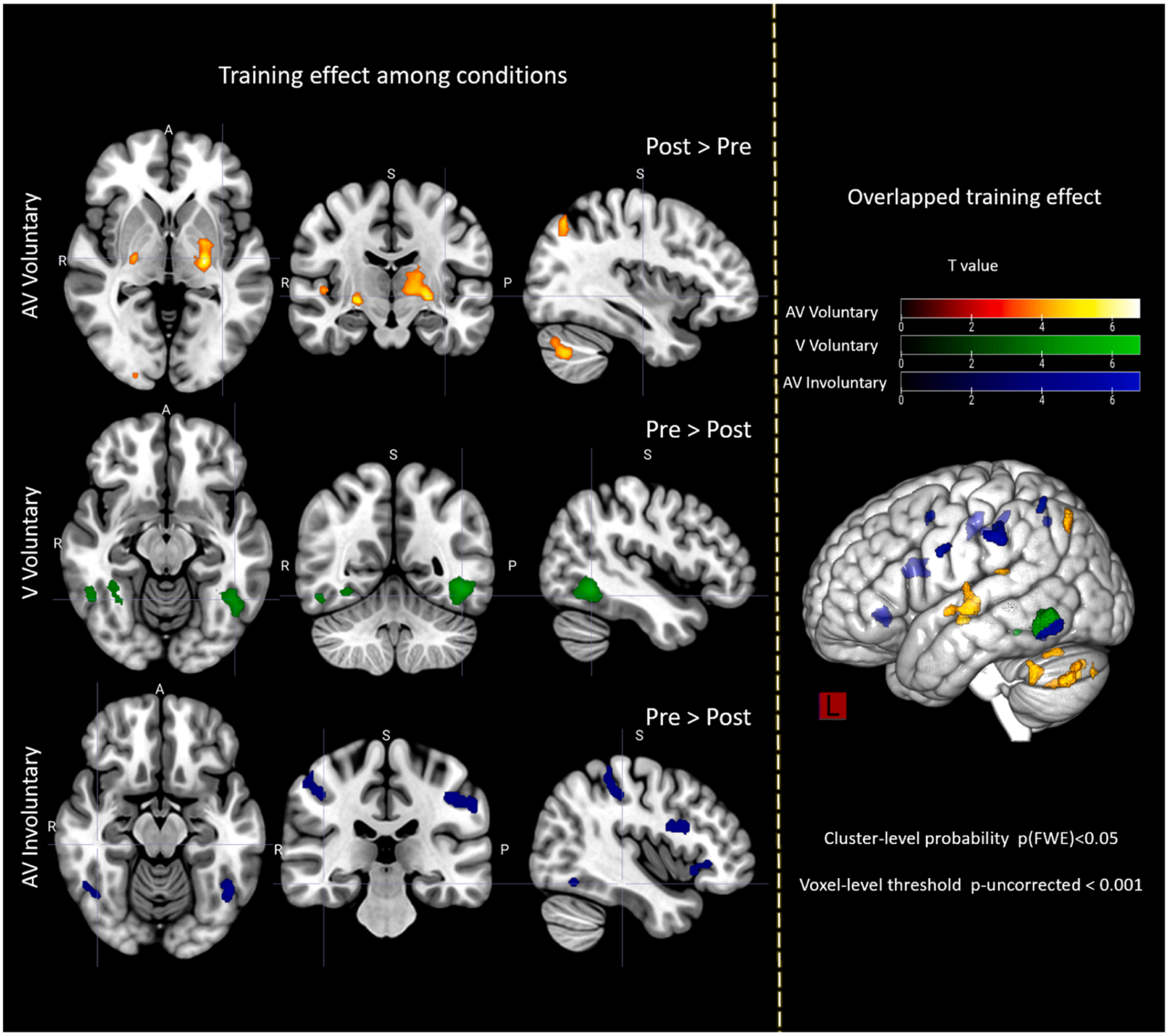

Fig. 6. Training effect functional activation: Significant activation changes between baseline and the end of training (post-training) in the three tasks (voxel-level threshold p uncorrected < 0.001): increase signal activation is seen in thalamus, cerebellum and inferior parietal lobe for the trained task in the top row (red), decrease signal activation for the V voluntary eye movement task in the inferior temporal gyrus in the middle row (green), and decrease signal activation for the AV involuntary eye movement task in the inferior temporal gyrus, supramarginal gyrus, and inferior frontal cortex in the bottom row (blue).

2.01; $p < 0.05$, $F_{(8152)}= 2.3$; $p < 0.05$, respectively), and significant main effects of 'condition' (IFC: $F_{(1.4,26.78)}= 17.5$; $p < 0.0001$, SG: $F_{(1.4, 26.56)}= 20.7$; $p < 0.0001$). A significant main effect of 'time' and 'condition' was found in the ITG (time: $F_{(4,76)} = 4.28$; $p < 0.01$, condition: $F_{(1.31, 24.96)} = 9.52$; $p < 0.01$). Post-hoc, Bonferroni corrected, paired *t*-tests showed that the activation decrease between baseline and post-training was significant in ITG only for the V voluntary eye movement task and was maintained in the follow-up session ($t_{(19)} =5.15$; $p < 0.0001$, $t_{(19)} =3.21$; $p < 0.05$, respectively, SupFig. 3). The functional activation during the AV involuntary task showed a significant decrease in the SG post-training ($t_{(19)} =3.35$; $p < 0.05$), which returned to baseline at the follow-up test ($t_{(19)} =1.75$; $p = 0.83$, SupFig. 3). No significant change was reported in any of the three regions for the AV voluntary (trained) eye movement task.

3.2.3. Correlation between fMRI and behavioural data

A Shapiro Test (Yap and Sim, 2011) was conducted to check normality for each measure, followed by a 2-tailed Pearson correlation test. This test revealed that the change in thalamus positively correlates with the change in behavioural data inside the MR scanner (increase in fMRI signal with increase of change in response time, $r = 0.61$; $p < 0.01$, Fig. 8a). No significant correlation between cerebellum, IPL, or ITG signal change and the behavioural measures were recorded. A negative correlation was observed between the activation change in the SG and IFC, and the response time measures (decrease in fMRI signal with increase of change in response time, $r=-69$; $p < 0.001$, $r=-53$; $p < 0.05$, Fig. 8b and c, respectively).

One-tailed Pearson correlation test showed a positive correlation between fMRI activation change in thalamus and cerebellum with the behavioural change in laboratory AV voluntary eye movement test (thalamus: $r = 0.44$; $p < 0.05$, cerebellum: $r = 0.4$; $p < 0.05$, SupFig. 4a and b). fMRI beta values at the baseline session correlate significantly with the behavioural change in thalamus and cerebellum but not in IPL ($r = 0.54$; $p < 0.05$ and $r = 0.49$; $p < 0.05$, respectively), which could predict the behavioural gains (SupFig. 5a and b). Baseline MRI activation in the thalamus and cerebellum is indicative of the potential for performance improvements during subsequent training, whereas this was not the case for the initial activation in the IPL.

Table 1

Brain regions table exported from SPM: this table shows the brain regions which showed significant signal activation change in the paired-T test for baseline and post training scans. All three tasks included (AV and V voluntary, and AV involuntary eye movement) with number of voxels for each brain region, statistical values, as well as the MNI coordinates.

Condition	Region	Number of voxels	Cluster-level	Peak-level	MNI coordinates(mm)		
		KE	p value (FWE)	T	X	Y	Z
Voluntary (AV)	Left Thalamus	337	0.001	7.18	−24	−18	−2
				4.69	−12	−16	2
				4.68	−24	−4	−2
	Right Cerebellum	185	0.013	5.44	34	−78	−36
				4.92	30	−82	−28
				4.57	34	−64	−40
	Left Cerebellum	200	0.01	5.2	−38	−68	−40
				4.46	−32	−74	−32
				4.18	−24	−72	−34
	Left Inferior Parietal lobe	121	0.037	4.89	−36	−70	50
				4.16	−36	−70	40
				4.09	−44	−72	40
Voluntary (V)	Left Inferior Temporal Gyrus	377	0.035	4.21	−46	−56	−8
				4.29	−42	−54	−16
				3.92	−36	−50	−10
Involuntary (AV)	Right Supramarginal Gyrus	429	0.003	6.15	56	−24	52
				4.57	40	−36	50
				4.33	46	−28	46
	Right Inferior Frontal Gyrus (Opercular part)	303	0.013	4.9	46	12	24
				4.52	40	4	22
				4.06	60	16	22
	Left Supramarginal Gyrus	266	0.023	4.86	−56	−32	38
				4.76	−48	−30	42
				4.68	−60	−24	38
	Right Inferior Frontal Gyrus (Orbital part)	248	0.03	4.85	36	30	0
				4.73	30	28	−6
				4.2	46	18	−10
	Left Inferior Temporal Gyrus	155	0.017	5.92	−48	−62	−10

4. Discussion

4.1. Behavioural improvement after AV voluntary eye movement training

The study's main objective was to explore the correlation between behavioural performance on a visuomotor learning task and functional parameters derived from neuroimaging when adding spatial and temporal congruent auditory cues. Consistent with prior work on VR training and AV learning (Seitz et al., 2006; Kim et al., 2008; Beer and Watanabe, 2009; Cooper et al., 2021; Valzolgher et al., 2020; Van den Brink et al., 2014), all participants exhibited significant improvements in behavioural performance over the course of the training, with a substantial reduction in mean response time and high response accuracy rates (> 90 %). This finding is consistent with Zou et al. (2012), who investigated the association between oculomotor scanning behaviour and the impact of non-spatial sounds. They proposed that non-spatial sounds elicit a general freezing effect on oculomotor scanning behaviour, thereby enhancing visual search performance through the facilitation of extended temporal and spatial information sampling. Additionally, Van den Brink et al. (2014) investigated behavioural effects of a concurrently presented auditory stimulus on visual target detection ability, found that individual's visual search efficiency was improved by the presence of the synchronized auditory signal (Van den Brink et al., 2014).

Our results, consistent with prior research (Van den Brink et al., 2014; Zou et al., 2012), demonstrate that incorporating auditory cues in the trained AV voluntary eye movement task significantly improved task performance, as evidenced by significantly faster response times in the audiovisual condition compared to the condition without audio cues. The findings are in line with previous studies discussed in a recent review (Alwashmi et al., 2022), which applied AV training for hemianopia patients and showed improvements in visual compensatory functions or restoration (Rowland et al., 2023). Our VR performance measures obtained during learning task follow an exponential decline with rapid gains (Cooper et al., 2021; Wang et al., 2004). Furthermore, our study showed that the participants' performance remained at the same level four weeks after the end of the training, indicating the persistent effect of the training as expected for motor and perceptual learning (Alwashmi et al., 2022; Aloufi et al., 2021).

VR has the capability of creating a motivating and interactive environment in which feedback can be provided and practice intensity can be monitored with minimal supervision. Feedback in the form of scores was given to motivate the participants to continuously improve their performance, which was influenced by both response times and accuracy rates, consistent with previous dual-task training studies (Cavaco et al., 2015; Bherer et al., 2005; Glass et al., 2000).

4.2. The transfer of learning to the untrained tasks

The study's results indicate that visual training accompanied by auditory cues can lead to learning, that transfers to improvements of voluntary eye movements even when the auditory cue is absent. The transfer of learning to the untrained tasks was, also, noticed, both involuntary and visual search tasks showed improvement after training in bimodal as well as unimodal conditions. A hallmark of the perceptual learning is that the improvements are strongly coupled to the trained stimulus and task, a characteristic that could limit its utility as a therapeutic tool. Early studies found transfer of learning to be highly specific to the trained eye and strongly monocular (Karni and Sagi, 1991), with no transfer between similar tasks (Fahle and Morgan, 1996). Yet, other studies suggest that sensory improvements derived from perceptual learning can transfer between different visual tasks (Zhang et al., 2010; McGovern et al., 2012). Other dual-task training programs have demonstrated that trained skills can transfer to other tasks that were not trained and retained for several months after training (Erickson et al., 2007; Bherer et al., 2005). This suggests that cognitive training programs can improve a general dual-task ability instead of strengthening a specific stimulus-response relationship for the trained task. However,

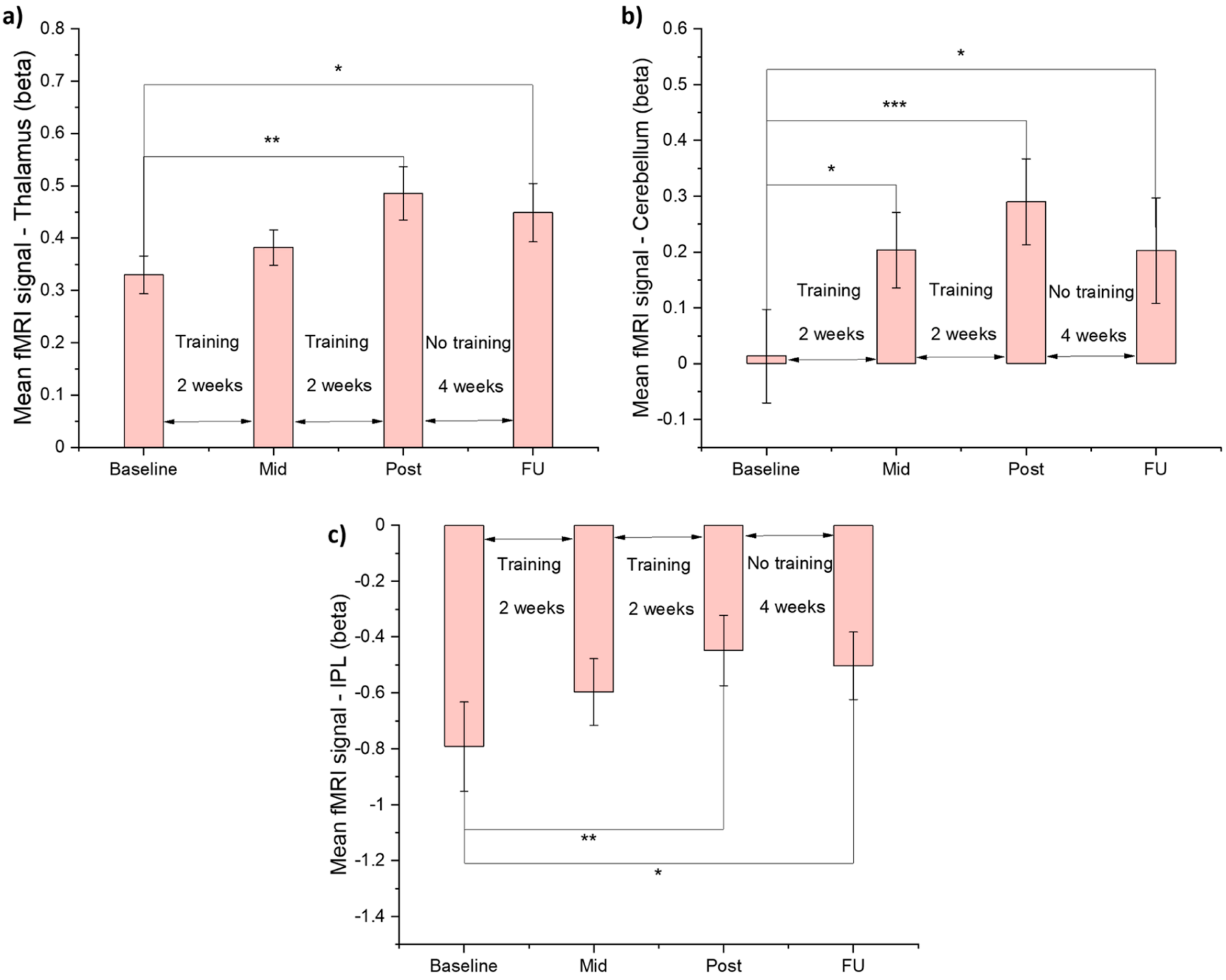

Fig. 7. Time course data extraction for functional measures from the areas that showed significant changes post-training for the trained task: the charts show the impact of the training on fMRI data for (a) the Thalamus, (b) the Cerebellum, and (c) the Inferior parietal lobe. The extracted imaging data represent, separately, the mean measures (fMRI beta) of all significant clusters in the AV voluntary eye movement task. fMRI metrics increased over the training period. However, while it reverts towards the baseline during the month after training ceases, it remains significantly different from the baseline. Error bars the standard error of the mean (SEM).

the transfer of learning is still a controversial topic in literature, which requires more investigation. The application of training on VR AV voluntary eye movement task and investigating its effect on untrained tasks with bimodal and unimodal conditions, combined with studying the underlying neural changes over time in our study may provide some explanations for the learning transfer process from the perspective of our training task.

4.3. *The interaction between multisensory integration and attention*

Attention is critical for selecting actions, but its role in eye movements is complex due to the fact that these movements can be either voluntary or involuntary, and in some situations, these two actions compete with each other, such as during visual search (DiGirolamo et al., 2016). Sounds can focus attention and facilitate response times for both voluntary and involuntary eye movements (DiGirolamo et al., 2016). Spatially and temporally congruent sound events with visual targets have been found to increase fixation durations when they occur and decrease the mean number of saccades in different eye movement strategies (Passamonti et al., 2009). Additionally, when the sound event is not in spatial coherence with the visual stimulus, but has a temporal congruency, it can facilitate the visual search mechanism. For instance, spatially uninformative sounds can help guide eye scanning away from previously viewed display regions that do not contain the target, improving search performance even on target-absent trials (Zou et al., 2012). These findings as well as what was previously mentioned in the introduction emphasise that AV integration requires attention, particularly when the auditory and visual stimuli are spatially and/or temporally coherent.

Three hypotheses have been proposed to describe the interaction between multisensory integration and attention (Talsma et al., 2010): (i) the stimulus-driven or bottom-up influence of multisensory integration on attention, whereby temporally aligned auditory input can affect the representation of co-occurring visual stimuli, making them stand out from competing stimuli (Van der Burg et al., 2008; Zou et al., 2012). (ii) the influence of top-down directed attention on multisensory integration, with spatial attention strongly affecting multiple stages of multisensory processing, indicating that spatial attention is coordinated across modalities, whether voluntarily or involuntarily directed, and the stimuli in different modalities do not have to be presented concurrently to be enhanced, but only need to occur in the spatially attended location (Talsma et al., 2010), and (iii) the spreading of attention across modalities, a process that combines aspects of both top-down and bottom-up mechanisms, occurring on a systems-level cascade of

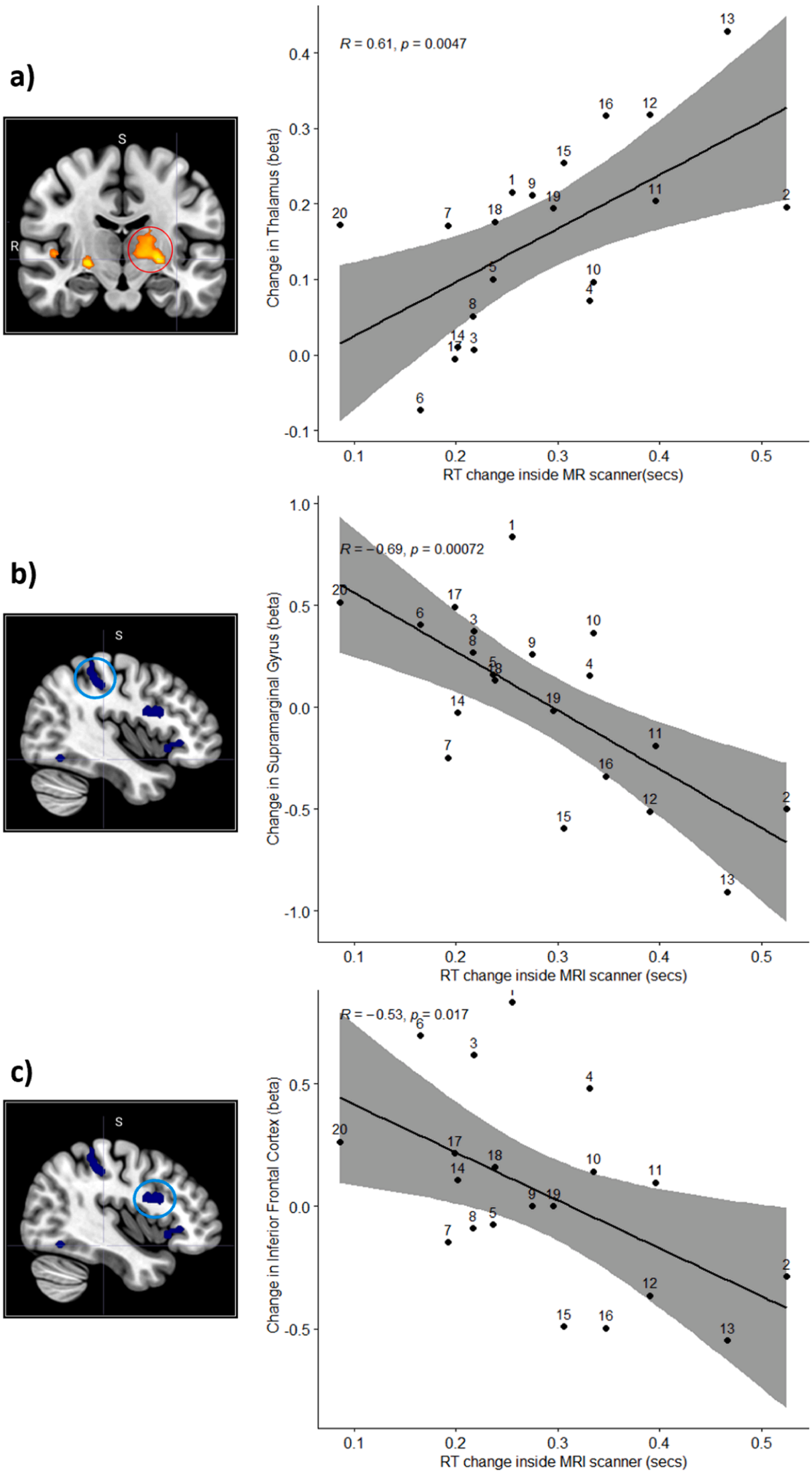

Fig. 8. Correlation between fMRI signal change and behavioural data: this figure shows a positive correlation between the fMRI signal increase of activation in the trained task and the change in behavioural data post-training inside the MR scanner in the thalamus (panel a), and a negative correlation between fMRI signal decrease of activation in the AV involuntary eye movement task with the behavioural change post-training inside the MR scanner in supramarginal gyrus (panel b) and inferior frontal gyrus (panel c).

top-down and bottom-up influences. This cascade starts with top-down visuospatial attention that selects which visual stimulus to process selectively, and then, through an automatic bottom-up linking mechanism resulting from multisensory components' temporal coincidence (Fleming et al., 2020; Stevenson et al., 2012; Talsma et al., 2010).

The spreading of attention may explain the improvement observed in both voluntary and involuntary tests caused by the temporal and spatial alignment of auditory cues in our results (Keller and Lefin-Rank, 2010; Grasso et al., 2016; Passamonti et al., 2009; Bolognini et al., 2005). This improvement is evidenced by an increase in accuracy and a decrease in search times. The improved efficiency and speed of visual scanning behaviour can be attributed to the stimulation of subcortical levels such as the superior colliculus (SC) and thalamus, which may promote the integration of sensory information and the implementation of more efficient oculomotor strategies (Froesel et al., 2021). The improvement found in both voluntary and involuntary tests in our study may be due to similar mechanisms of AV integration and oculomotor strategy implementation. To this end, the interpretation of the neuroimaging data can help answer the important questions of how the auditory cues facilitate the oculomotor process, how the transfer of learning could happen, and what brain regions are involved in these processes.

4.4. *Neuroplastic changes*

4.4.1. *Functional activation during AV voluntary eye movement*

Most sensory information reaches the neocortex through the superior colliculus and thalamus (Noesselt et al., 2010; Froesel et al., 2021). The pulvinar, the largest and most posterior thalamic nucleus, has strong feedforward and feedback connections with the cortex as well as with the superior colliculus (Froesel et al., 2021). Spatial deficits in localising visual targets were reported following damage to the human thalamus (pulvinar) (Ward and Arend, 2007). Previous neuroimaging studies (Tyll et al., 2011; Bonath et al., 2013; Noesselt et al., 2010; Van den Brink et al., 2014; Baier et al., 2006; Komura et al., 2005) showed the important role of the thalamus in multisensory integration both in humans and animals. Komura et al. (2005) found that visual cues in the rat thalamus influence auditory responses, demonstrating that cross-modal cueing modulates gain in the sensory thalamus, potentially providing a priming influence on the choice of optimal behaviour. Increase functional activation in the thalamus has been reported by Bonath et al. (2013) after AV training on spatial and temporal congruency judgment (Bonath et al., 2013). Noesselt et al. (2010) tested whether a co-occurring sound could enhance visual target detection sensitivity and depicted functional activation in the thalamus when combining visual and auditory stimulation. These results are in line with our finding of the change we found in the thalamus after training.

The inferior parietal lobe (including supramarginal gyrus and angular gyrus), inferior frontal cortex, and superior temporal cortex are known to play important roles in visual-spatial localisation, spatial attention, and cognition, with damage to these regions resulting in spatial neglect (Perry and Zeki, 2000; Shapiro et al., 2002; Lunven and Bartolomeo, 2017; Meyer et al., 2013). Previous research has shown that a common network in the parietal, frontal, and temporal lobes is activated during tasks involving both attentional and saccadic shifts to peripheral visual stimuli, indicating tight integration between attentional and oculomotor processes (Corbetta et al., 1998).

One fMRI study by Bolger et al. (2014) found that auditory and visual detection tasks involving metrical rhythm with its relatively strong and weak beats modulated attentional resources over time, leading to increased signal activation in the left IPL. Furthermore, two other studies have reported heightened activation in the inferior frontal gyrus and left inferior parietal cortex during bimodal and trimodal stimulation (olfactory, visual, and auditory) (Stickel et al., 2019; Porada et al., 2021).

These findings suggest that multisensory integration plays an important role in attentional processes and that the left IPL may be particularly important in processing auditory and visual information in an oculomotor context. Our task involves both components of attention (auditory cues) and oculomotor functions (systematic eye movement), and we found an increase in activation in the inferior parietal lobe, which is in line with these previous findings.

Upon closer examination of our findings, we observed that the inferior parietal lobe (IPL) initially showed a deactivation during the performance of the voluntary AV task at baseline, but this deactivation decreased significantly over time as evidenced by the extracted beta values. Mofrad and Schiller (2022) investigated functional connectivity during a cognitive control task, found negative functional associations between the left caudal IPL and the lateral occipital cortex, indicating enhanced cognitive performance. The authors suggest that the absence of the positive functional connectivity effect reported in the literature between the IPL and visual regions could be due to the deactivation of the caudal IPC during visual task performance (Mofrad and Schiller, 2022). Nevertheless, the decrease in deactivation we found may be attributed to the visuospatial function of the IPL, which is recruited through the learning of the congruency between the auditory and visual stimuli to achieve faster response times in our trained AV voluntary task.

It's worth noting that, in our investigation, we reported alterations within the IPL, which may overlap with the IPS depending on the MNI coordinates, as it has been previously characterized as both IPL and IPS in existing literature (Matejko et al., 2019). Recent findings indicate that patients afflicted with lesions in the right inferior parietal lobule (IPL), including the right intraparietal sulcus (IPS) and the underlying white matter, exhibit deficiencies in sustaining attention during spatial tasks (Malhotra et al., 2009). A study conducted by Lee et al. (2013) delved further into the roles of the left and right IPL, after excluding the IPS, and explored sustained attention using transcranial magnetic stimulation (TMS) to emulate virtual lesions in the left and right superior parietal lobe (SPL) and IPL. Their investigation encompassed two distinct categories of visual sustained attention tasks: spatial (location-based) and non-spatial (feature-based). During the spatial task, repetitive (rTMS) applied over either the right or left IPL resulted in a decline in sustained attention, manifesting as a progressive increase in errors and response times among participants. In contrast, rTMS had no discernible impact on participants' performance in the non-spatial task (Lee et al., 2013). These observations bear relevance to the findings of our own study, where the alignment of spatial and temporal cues across distinct sensory modalities appeared to exert a substantial influence on selection and processing of our attention.

The presence of fMRI activation changes in both the thalamus and caudal inferior parietal cortex following audio-visual training could be considered evidence for the involvement of both top-down and bottom-up attentional mechanisms. The thalamus is known to be involved in both top-down and bottom-up attentional processes (Tokoro et al., 2015). Changes in fMRI activation within the thalamus could suggest the modulation of attentional control and the regulation of sensory inputs because of the temporal alignment, reflecting top-down mechanism (Tokoro et al., 2015). The caudal inferior parietal cortex, on the other hand, is often associated with the processing of sensory inputs and the integration of multisensory information (Ruschel et al., 2014). This region plays a crucial role in bottom-up attentional processes, where the alignment of spatial cues from different sensory modalities can influence attentional selection and processing (Ruschel et al., 2014). Changes in fMRI activation within the caudal inferior parietal cortex after audio-visual training could indicate the reorganization or enhancement of sensory integration processes.

In addition, studies have highlighted the cerebellum's significant contribution to multisensory perceptual learning, suggesting that its activation is linked to the perceptual requirements of a task (Baumann and Mattingley, 2010; Baumann et al., 2015). Specifically, activity in the left crus I region was significantly correlated with increases in perceptual demands for both auditory and visual stimuli, indicating a critical role of this region in multisensory processing (Baumann and Mattingley, 2010). In addition, Baumann et al. (2015) found that AV motion detection led to increased activity in cerebellar lobule VI and right lateral crus I, compared to unimodal visual and auditory motion tasks, further emphasising the cerebellum's involvement in multisensory integration. Consistent with previous findings, our study revealed an increase in fMRI signal activation in both the right and left cerebellar crus I regions.

In our discussion, we have emphasized the brain regions that exhibited significant changes in activation following our training protocol, noting the absence of any discernible alterations in the superior temporal sulcus (STS) post-training which widely known to be linked with audio-visual integration. A comprehensive review conducted by Hein and Knight (2008) sheds light on the typical activations observed in the STS across a multitude of fMRI studies. However, it is worth noting that investigations pertaining to AV integration in these studies predominantly employed semantic categorization or semantic congruency stimuli such as images and sounds of common objects (Taylor et al., 2006; Hein et al., 2007) or written and spoken letters compared to fixation (Van Atteveldt et al., 2004; Blau et al., 2008). Additionally, Meyer et al. (2011) showed semantically incongruent stimuli, consisting of two meaningful speech and body action signals that were presented simultaneously in different modalities, causes increased activation in posterior temporal (pSTS). These incongruency effects were not seen when a meaningful signal was paired with a meaningless signal (Meyer et al., 2011). This suggests that these overlapping networks are used to process meaningful speech and body-action signals. These AV paradigms differ significantly from the AV spatial localization task we employed in our study.

4.4.2. Functional activation during AV involuntary task

As noted previously, the spreading of attention hypothesis could provide an explanation for the learning transfer we observed in the involuntary task following training on the voluntary task. Consequently, we aimed to investigate the neural changes associated with this process. A previous study by Aloufi et al. (2021) employed a similar training paradigm to ours and found no noticeable changes in behaviour or brain function in the involuntary task, after six weeks of training on a unimodal voluntary eye movement task. This result suggests that the changes we observed in fMRI activation during the involuntary task may have resulted from the activation of multisensory areas through the use of auditory cues, which are involved in the visuospatial attentional process.

Our study found a significant decrease in fMRI activation in the IFG and SG while performing the AV involuntary task after training. Previous studies have linked the decrease in fMRI activation to improved performance, indicating that the practice could decrease the reliance on attentional resources and lead to increased neural efficiency (Erickson et al., 2007; Tomasi et al., 2004). This suggests that as participants became more adept at the task, there was a transition from a state that required more neural resources to perform the novel task to more automated processing, leading to a reduction in fMRI signal (Erickson et al., 2007; Tomasi et al., 2004).

The inferior frontal cortex and supramarginal gyrus of the inferior parietal lobe play important roles in spatial attention as part of the attentional network. Krumbholz et al. (2009) explored whether the same or different brain areas are responsible for auditory and visual-spatial attention and found that there is some overlap between the two conditions in multiple brain regions, including the inferior frontal and parietal cortices. Another study by Umarova et al. (2010) focused on the visual spatial attention task and revealed that the inferior parietal and inferior frontal cortices were both activated and involved in the attentional system. Additionally, the superior longitudinal fasciculus II and III and the arcuate fasciculus, which are white matter tracts that travel within the white matter of the inferior parietal lobe, provide structural connectivity for spatial attention by connecting the temporoparietal regions with the inferior frontal cortex and the frontal eye field (Umarova et al., 2010). These findings could elucidate the direct connection between the IPL and the IFG required for visuospatial attention, which was shown to have activation changes after training in our results.

4.4.3. Functional activation during unimodal visual voluntary eye movement

As part of this study, we aimed to investigate the effect of multisensory training on voluntary eye movements without auditory cues. Our findings showed a significant decrease in functional activation in the inferior temporal gyrus (ITG), which has been previously reported during visual skill learning (Poldrack et al., 1998). The shift found in neural processing from superior parietal and posterior occipital to inferior temporal cortices during skill learning may reflect the concept of increased neural efficiency after training, which has previously explained at the initial exposure to a novel task (Tomasi et al., 2004). The ITG has been suggested to play a vital role in the visual processing of orthographic information as part of the ventral occipitotemporal stream, including posterior and anterior fusiform, inferior temporal, and lateral occipital regions (Conway, 2018; Quinn et al., 2017). During the initial learning stage, the AV representation is encoded, and short-term memory of the AV objects is stored in the middle and inferior temporal regions, and possibly also in the medial temporal regions such as the hippocampus (Xu et al., 2020). This region is essential in our task as it is required to recognise the stimuli' shape and respond as a matter of accurate quantification. In addition, ITG activation was reported during a congruent AV judgment task as a part of the higher visual areas involved in multisensory integration (Meienbrock et al., 2007). Therefore, the reduction in activation in the ITG during the AV involuntary task as well as the unimodal voluntary task might be an indirect effect of our congruent AV training task.

4.5. Linking neuroplastic to behavioural change

In this study, a notable finding was the significant correlation between changes in functional activation and behavioural performance changes in both AV voluntary and involuntary tasks, suggesting that the observed neuroplastic changes in the target areas support performance improvements (Plank et al., 2012). Specifically, increases in fMRI signal activation in the thalamus and cerebellum post-training during the voluntary eye movement task, and decreases in activation in the SG and IFC during the involuntary task were significantly correlated with improved performance. Previous studies have also reported correlations between changes in brain activations and performance improvement during training on a contrast discrimination task (Mukai et al., 2007) and systematic voluntary eye movement task (Aloufi et al., 2021). Furthermore, the study's correlation test revealed that participants with higher fMRI signal activation at baseline in the cerebellum and thalamus showed greater improvement in behavioural performance. These findings suggest that initial task-specific brain activation patterns may be used to predict whether a subject is likely to improve with training (Mukai et al., 2007). Overall, the correlation between fMRI activation and behavioural improvement supports the idea that the observed changes in these brain regions are a result of learning.

5. Conclusion

In summary, this study investigated the neural mechanisms underlying multisensory learning using an AV voluntary eye movement task in a virtual environment. The results revealed changes in functional activation in various brain regions where multisensory integration occurs in cortical and subcortical levels, including the thalamus, IPL, ITG, SG, IFC, and cerebellum, which were associated with the improvement in behavioural performance. Spreading of attention might explain the learning transfer after training on the AV voluntary eye movement to the AV and V involuntary and visual search tasks caused by the temporal and spatial alignment of auditory cues. These findings suggest that multisensory learning involves changes in the processing of sensory information, neural efficiency, and functional connectivity between different brain regions. Moreover, the study provides evidence that training-induced neuroplastic changes in these areas support behavioural improvement, and the baseline fMRI signal activation in task-specific areas during the initial period of training may predict the learning outcomes. Depending on our findings and previous research, we suggest that visual deficit rehabilitation could benefit from the application of multisensory training in VR, as the overlap between the attentional and oculomotor processes, which is required for multisensory integration may provide different neural circuits that may support the visual compensatory strategy, however, this requires more investigation in future research. In addition, these findings could have important implications for the development of effective training strategies in more general visual skill learning.

Data and code availability statement

All data and code are available from the authors upon request.

CRediT authorship contribution statement

Kholoud Alwashmi: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Georg Meyer:** Conceptualization, Investigation, Methodology, Data curation, Resources, Project administration, Supervision, Validation, Visualization, Writing – review & editing. **Fiona Rowe:** Conceptualization, Methodology, Resources, Supervision, Validation, Visualization, Writing – review & editing. **Ryan Ward:** Methodology, Visualization, Writing – original draft.

Declaration of Competing Interest

The authors report no relevant financial or non-financial conflicts in relation to this study.

Data availability

Data will be made available on request.

Acknowledgments and funding sources

This research was funded by Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia, as the first author is a sponsored PhD student. We thank our lab members from the University of Liverpool especially Prof. Sophie Wuerger, and Dr. Sahal Alotaibi, who provided insight and expertise that greatly assisted the research.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2023.120483.

References

Allue, M., Serrano, A., Bedia, M.G., Masia, B., 2016. Crossmodal perception in immersive environments. In: Proceedings of the XXVI Spanish Computer Graphics Conference, pp. 1–7.

Aloufi, A.E., Rowe, F.J., Meyer, G.F., 2021. Behavioural performance improvement in visuomotor learning correlates with functional and microstructural brain changes. Neuroimage 227, 117673.

Altieri, N., Stevenson, R.A., Wallace, M.T., Wenger, M.J., 2015. Learning to associate auditory and visual stimuli: behavioral and neural mechanisms. Brain Topogr. 28, 479–493.

Alwashmi, K., Meyer, G., Rowe, F.J., 2022. Audio-visual stimulation for visual compensatory functions in stroke survivors with visual field defect: a systematic review. Neurological Sciences 43 (4), 2299–2321.

Anzanello, M.J., Fogliatto, F.S., 2011. Learning curve models and applications: literature review and research directions. Int. J. Ind. Ergon. 41 (5), 573–583.

Baier, B., Kleinschmidt, A., Müller, N.G., 2006. Cross-modal processing in early visual and auditory cortices depends on expected statistical relationship of multisensory information. J. Neurosci. 26 (47), 12260–12265.

Baumann, O., Mattingley, J.B., 2010. Scaling of neural responses to visual and auditory motion in the human cerebellum. J. Neurosci. 30 (12), 4489–4495.

Baumann, O., et al., 2015. Consensus paper: the role of the cerebellum in perceptual processes. Cerebellum 14, 197–220.

Beer, A.L., Watanabe, T., 2009. Specificity of auditory-guided visual perceptual learning suggests crossmodal plasticity in early visual cortex. Exp. Brain Res. 198, 353–361.

Beer, A.L., Batson, M.A., Watanabe, T., 2011. Multisensory perceptual learning reshapes both fast and slow mechanisms of crossmodal processing. Cogn. Affect. Behav. Neurosci. 11 (1), 1–12.

Beig, M., Kapralos, B., Collins, K., Mirza-Babaei, P., 2019. An introduction to spatial sound rendering in virtual environments and games. Comput. Games J. 8 (3–4), 199–214.

Bernstein, I.H., Edelstein, B.A., 1971. Effects of some variations in auditory input upon visual choice reaction time. J. Exp. Psychol. 87 (2), 241.

Bherer, L., Kramer, A.F., Peterson, M.S., Colcombe, S., Erickson, K., Becic, E., 2005. Training effects on dual-task performance: are there age-related differences in plasticity of attentional control? Psychol. Aging 20 (4), 695.

Blau, V., Van Atteveldt, N., Formisano, E., Goebel, R., Blomert, L., 2008. Task-irrelevant visual letters interact with the processing of speech sounds in heteromodal and unimodal cortex. Eur. J. Neurosci. 28 (3), 500–509.

Bolger, D., Coull, J.T., Schön, D., 2014. Metrical rhythm implicitly orients attention in time as indexed by improved target detection and left inferior parietal activation. J. Cogn. Neurosci. 26 (3), 593–605.

Bolognini, N., Rasi, F., Coccia, M., Ladavas, E., 2005. Visual search improvement in hemianopic patients after audio-visual stimulation. Brain 128 (12), 2830–2842.

Bonath, B., Tyll, S., Budinger, E., Krauel, K., Hopf, J.M., Noesselt, T., 2013. Task-demands and audio-visual stimulus configurations modulate neural activity in the human thalamus. Neuroimage 66, 110–118.

Brett, M., Anton, J.L., Valabregue, R., Poline, J.B., 2002. Region of interest analysis using an SPM toolbox. In: Proceedings of the 8th International Conference on Functional Mapping of the Human Brain. Sendai, 16, p. 497.

Cavaco, S., Simões, D., Silva, T., 2015. Spatialized audio in a vision rehabilitation game for training orientation and mobility skills. In: Proceedings of the International Conference on Digital Audio E ects (DAFx).

Chiou, R., Rich, A.N., 2012. Cross-modality correspondence between pitch and spatial location modulates attentional orienting. Perception 41 (3), 339–353.

Conway, B.R., 2018. The organization and operation of inferior temporal cortex. Annu. Rev. Vis. Sci. 4, 381–402.

Cooper, N., Milella, F., Pinto, C., Cant, I., White, M., Meyer, G., 2018. The effects of substitute multisensory feedback on task performance and the sense of presence in a virtual reality environment. PLoS One 13 (2), e0191846.

Cooper, N., Millela, F., Cant, I., White, M.D., Meyer, G., 2021. Transfer of training—Virtual reality training with augmented multisensory cues improves user experience during training and task performance in the real world. PLoS One 16 (3), e0248225.

Corbetta, M., et al., 1998. A common network of functional areas for attention and eye movements. Neuron 21 (4), 761–773.

Danevičius, E., Stief, F., Matynia, K., Larsen, M.L., Kraus, M., 2021. 3D localisation of sound sources in virtual reality. In: Proceedings of the Interactivity and Game Creation: 9th EAI International Conference, ArtsIT 2020. Aalborg, Denmark, 9. Springer, pp. 307–319. *December 10–11, 2020*, Proceedings.

DiGirolamo, G.J., Patel, N., Blaukopf, C.L., 2016. Arousal facilitates involuntary eye movements. Exp. Brain Res. 234, 1967–1976.

Driver, J., Noesselt, T., 2008. Multisensory interplay reveals crossmodal influences on 'sensory-specific'brain regions, neural responses, and judgments. Neuron 57 (1), 11–23.

Dundon, N.M., Bertini, C., Làdavas, E., Sabel, B.A., Gall, C., 2015. Visual rehabilitation: visual scanning, multisensory stimulation and vision restoration trainings. Front. Behav. Neurosci. 9, 192.

Erickson, K.I., et al., 2007. Training-induced functional activation changes in dual-task processing: an FMRI study. Cereb. Cortex 17 (1), 192–204.

Evans, K.K., Treisman, A., 2010. Natural cross-modal mappings between visual and auditory features. J. Vis. 10 (1), 6. -6.

Fahle, M., Morgan, M., 1996. No transfer of perceptual learning between similar stimuli in the same retinal position. Curr. Biol. 6 (3), 292–297.

Fleming, J.T., Noyce, A.L., Shinn-Cunningham, B.G., 2020. Audio-visual spatial alignment improves integration in the presence of a competing audio-visual stimulus. Neuropsychologia 146, 107530.

fMRI solution "NordicNeuroLab". 2023. [Online]. Available at: https://www.nordicneurolab.com/.

Frassinetti, F., Bolognini, N., Làdavas, E., 2002. Enhancement of visual perception by crossmodal visuo-auditory interaction. Exp. Brain Res. 147 (3), 332–343.

Friston, K.J., Holmes, A.P., Price, C., Büchel, C., Worsley, K., 1999. Multisubject fMRI studies and conjunction analyses. Neuroimage 10 (4), 385–396.

Froesel, M., Cappe, C., Hamed, S.B., 2021. A multisensory perspective onto primate pulvinar functions. Neurosci. Biobehav. Rev. 125, 231–243.

Furmanski, C.S., Schluppeck, D., Engel, S.A., 2004. Learning strengthens the response of primary visual cortex to simple patterns. Curr. Biol. 14 (7), 573–578.

Ghazanfar, A.A., Schroeder, C.E., 2006. Is neocortex essentially multisensory? Trends Cogn. Sci. 10 (6), 278–285.

Glass, J.M., et al., 2000. Aging and the psychological refractory period: task-coordination strategies in young and old adults. Psychol. Aging 15 (4), 571.

Grasso, P.A., Làdavas, E., Bertini, C., 2016. Compensatory recovery after multisensory stimulation in hemianopic patients: behavioral and neurophysiological components. Front. Syst. Neurosci. 10, 45.

Greenhouse, S.W., Geisser, S., 1959. On methods in the analysis of profile data. Psychometrika 24 (2), 95–112.

Hein, G., Knight, R.T., 2008. Superior temporal sulcus—It's my area: or is it? J. Cogn. Neurosci. 20 (12), 2125–2136.

Hein, G., Doehrmann, O., Müller, N.G., Kaiser, J., Muckli, L., Naumer, M.J., 2007. Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. J. Neurosci. 27 (30), 7881–7887.

Hofman, P.M., Van Riswick, J.G., Van Opstal, A.J., 1998. Relearning sound localization with new ears. Nat. Neurosci. 1 (5), 417–421.

Huygelier, H., et al., 2022. An immersive virtual reality game to train spatial attention orientation after stroke: a feasibility study. Appl. Neuropsychol. Adult 29 (5), 915–935.

Karni, A., Sagi, D., 1991. Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. Proc. Natl. Acad. Sci. 88 (11), 4966–4970.

Keller, I., Lefin-Rank, G., 2010. Improvement of visual search after audiovisual exploration training in hemianopic patients. Neurorehabil. Neural Repair 24 (7), 666–673.

Kim, R.S., Seitz, A.R., Shams, L., 2008. Benefits of stimulus congruency for multisensory facilitation of visual learning. PLoS One 3 (1), e1532.

Koelewijn, T., Bronkhorst, A., Theeuwes, J., 2010. Attention and the multiple stages of multisensory integration: a review of audiovisual studies. Acta Psychol. 134 (3), 372–384.

Komura, Y., Tamura, R., Uwano, T., Nishijo, H., Ono, T., 2005. Auditory thalamus integrates visual inputs into behavioral gains. Nat. Neurosci. 8 (9), 1203–1209.

Kourtzi, Z., Betts, L.R., Sarkheil, P., Welchman, A.E., 2005. Distributed neural plasticity for shape learning in the human visual cortex. PLoS Biol. 3 (7), e204.

Krumbholz, K., Nobis, E.A., Weatheritt, R.J., Fink, G.R., 2009. Executive control of spatial attention shifts in the auditory compared to the visual modality. Hum. Brain Mapp. 30 (5), 1457–1469.

Larsen, C.R., et al., 2009. Effect of virtual reality training on laparoscopic surgery: randomised controlled trial. BMJ 338.

Lauzon, S.Á., Abraham, A.E., Curcin, K., Butler, B.E., Stevenson, R.A., 2022. The relationship between multisensory associative learning and multisensory integration. Neuropsychologia 174, 108336.

Lee, J., et al., 2013. rTMS over bilateral inferior parietal cortex induces decrement of spatial sustained attention. Front. Hum. Neurosci. 7, 26.

Lu, Z.L., Hua, T., Huang, C.B., Zhou, Y., Dosher, B.A., 2011. Visual perceptual learning. Neurobiol. Learn. Mem. 95 (2), 145–151.

Lunven, M., Bartolomeo, P., 2017. Attention and spatial cognition: neural and anatomical substrates of visual neglect. Ann. Phys. Rehabil. Med. 60 (3), 124–129.

Macaluso, E., Driver, J., 2005. Multisensory spatial interactions: a window onto functional integration in the human brain. Trends Neurosci. 28 (5), 264–271. https://doi.org/10.1016/j.tins.2005.03.008, 2005/05/01/.

Malhotra, P., Coulthard, E.J., Husain, M., 2009. Role of right posterior parietal cortex in maintaining attention to spatial locations over time. Brain 132 (3), 645–660.

Malpica, S., Serrano, A., Allue, M., Bedia, M.G., Masia, B., 2020. Crossmodal perception in virtual reality. Multimed. Tools Appl. 79 (5), 3311–3331.

Mannan, S.K., Pambakian, A.L., Kennard, C., 2010. Compensatory strategies following visual search training in patients with homonymous hemianopia: an eye movement study. J. Neurol. 257 (11), 1812–1821.

Matejko, A.A., Hutchison, J.E., Ansari, D., 2019. Developmental specialization of the left intraparietal sulcus for symbolic ordinal processing. Cortex 114, 41–53.

MATLAB, "Detect and remove outliers in data - MATLAB rmoutliers - MathWorks United Kingdom," 2023. [Online]. Available at: https://uk.mathworks.com/help/matlab/ref/rmoutliers.html.

Maus, B., Gerard, J., van Breukelen, P., Goebel, R., Martijn, P., Berger, F., 2010. Optimization of blocked designs in fMRI studies. Psychometrika 75 (2), 373.

McCormick, K., Lacey, S., Stilla, R., Nygaard, L.C., Sathian, K., 2018. Neural basis of the crossmodal correspondence between auditory pitch and visuospatial elevation. Neuropsychologia 112, 19–30. https://doi.org/10.1016/j.neuropsychologia.2018.02.029, 2018/04/01/.

McGovern, D.P., Webb, B.S., Peirce, J.W., 2012. Transfer of perceptual learning between different visual tasks. J. Vis. 12 (11), 4. -4.

Meienbrock, A., Naumer, M., Doehrmann, O., Singer, W., Muckli, L., 2007. Retinotopic effects during spatial audio-visual integration. Neuropsychologia 45 (3), 531–539.

Meta Quest 2 "Immersive All-In-One VR Headset". 2023. [Online]. Available at: https://www.meta.com/quest/products/quest-2/.

Meyer, G.F., Wuerger, S., 2001. Cross-modal integration of auditory and visual motion signals. Neuroreport 12 (11), 2557–2560.

Meyer, G.F., Wuerger, S.M., Röhrbein, F., Zetzsche, C., 2005. Low-level integration of auditory and visual motion signals requires spatial co-localisation. Exp. Brain Res. 166, 538–547.

Meyer, G.F., Greenlee, M., Wuerger, S., 2011. Interactions between auditory and visual semantic stimulus classes: evidence for common processing networks for speech and body actions. J. Cogn. Neurosci. 23 (9), 2291–2308.

Meyer, G.F., Harrison, N.R., Wuerger, S.M., 2013. The time course of auditory–visual processing of speech and body actions: evidence for the simultaneous activation of an extended neural network for semantic processing. Neuropsychologia 51 (9), 1716–1725.

Mofrad, F.T., Schiller, N.O., 2022. Mapping caudal inferior parietal cortex supports the hypothesis about a modulating cortical area. Neuroimage 259, 119441.

Mudd, S., 1963. Spatial stereotypes of four dimensions of pure tone. J. Exp. Psychol. 66 (4), 347.

Mukai, I., Kim, D., Fukunaga, M., Japee, S., Marrett, S., Ungerleider, L.G., 2007. Activations in visual and attention-related areas predict and correlate with the degree of perceptual learning. J. Neurosci. 27 (42), 11401–11411.

Murray, M.M., Thelen, A., Thut, G., Romei, V., Martuzzi, R., Matusz, P.J., 2016. The multisensory function of the human primary visual cortex. Neuropsychologia 83, 161–169.

Noesselt, T., Tyll, S., Boehler, C.N., Budinger, E., Heinze, H.J., Driver, J., 2010. Sound-induced enhancement of low-intensity vision: multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity. J. Neurosci. 30 (41), 13609–13623.

Oculus developers l MetaQuest "Localization and the human auditory system". 2023. [Online]. Available at: https://developer.oculus.com/resources/audio-intro-localization/.

Passamonti, C., Bertini, C., Làdavas, E., 2009. Audio-visual stimulation improves oculomotor patterns in patients with hemianopia. Neuropsychologia 47 (2), 546–555.

Peirce, J., et al., 2019a. PsychoPy2: experiments in behavior made easy. Behav. Res. Methods 51 (1), 195–203.

Peirce, J., et al., 2019b. PsychoPy2: experiments in behavior made easy. Behav. Res. Methods 51, 195–203.

Perry, R., Zeki, S., 2000. The neurology of saccades and covert shifts in spatial attention: an event-related fMRI study. Brain 123 (11), 2273–2288.

Plank, T., Rosengarth, K., Song, W., Ellermeier, W., Greenlee, M.W., 2012. Neural correlates of audio-visual object recognition: effects of implicit spatial congruency. Hum. Brain Mapp. 33 (4), 797–811.

Poldrack, R.A., Desmond, J.E., Glover, G.H., Gabrieli, J., 1998. The neural basis of visual skill learning: an fMRI study of mirror reading. Cereb. Cortex 8 (1), 1–10 (New York, NY: 1991).

Porada, D.K., Regenbogen, C., Freiherr, J., Seubert, J., Lundström, J.N., 2021. Trimodal processing of complex stimuli in inferior parietal cortex is modality-independent. Cortex 139, 198–210.

Powers, A.R., Hillock, A.R., Wallace, M.T., 2009. Perceptual training narrows the temporal window of multisensory binding. J. Neurosci. 29 (39), 12265–12274.

Powers, A.R., Hevey, M.A., Wallace, M.T., 2012. Neural correlates of multisensory perceptual learning. J. Neurosci. 32 (18), 6263–6274.

Pratt, C.C., 1930. The spatial character of high and low tones. J. Exp. Psychol. 13 (3), 278.

Quinn, C., Taylor, J., Davis, M.H., 2017. Learning and retrieving holistic and componential visual-verbal associations in reading and object naming. Neuropsychologia 98, 68–84.

REX, "Rex manual 5 1 09 - MIT". 2023 [Online]. Available at: https://web.mit.edu/swg/rex/rex.pdf.

Risoud, M., et al., 2018. Sound source localization. Eur. Ann. Otorhinolaryngol. Head Neck Dis. 135 (4), 259–264.

Roffler, S.K., Butler, R.A., 1968. Factors that influence the localization of sound in the vertical plane. J. Acoust. Soc. Am. 43 (6), 1255–1259.

Rowe, F.J., et al., 2017. A pilot randomized controlled trial comparing effectiveness of prism glasses, visual search training and standard care in hemianopia. Acta Neurol. Scand. 136 (4), 310–321.

Rowland, B.A., Bushnell, C.D., Duncan, P.W., Stein, B.E., 2023. Ameliorating hemianopia with multisensory training. J. Neurosci. 43 (6), 1018–1026.

Ruschel, M., Knösche, T.R., Friederici, A.D., Turner, R., Geyer, S., Anwander, A., 2014. Connectivity architecture and subdivision of the human inferior parietal cortex revealed by diffusion MRI. Cereb. Cortex 24 (9), 2436–2448.

Sürig, R., Bottari, D., Röder, B., 2018. Transfer of audio-visual temporal training to temporal and spatial audio-visual tasks. Multisens. Res. 31 (6), 556–578.

Sasaki, Y., Nanez, J.E., Watanabe, T., 2010. Advances in visual perceptual learning and plasticity. Nat. Rev. Neurosci. 11 (1), 53–60.

Seitz, A.R., Kim, R., Shams, L., 2006. Sound facilitates visual learning. Curr. Biol. 16 (14), 1422–1427.

Seymour, N.E., et al., 2002. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann. Surg. 236 (4), 458.

Shapiro, K., Hillstrom, A.P., Husain, M., 2002. Control of visuotemporal attention by inferior parietal and superior temporal cortex. Curr. Biol. 12 (15), 1320–1325.

Singh, V., Rana, R.K., Singhal, R., 2013. Analysis of repeated measurement data in the clinical trials. J. Ayurveda Integr. Med. 4 (2), 77.

SPM 12 "statistical parametric mapping". 2023 Functional Imaging Laboratory.[Online] Available at: https://www.fil.ion.ucl.ac.uk/spm/.

Stein, B.E., Meredith, M.A., 1993. The Merging of the Senses. MIT press.

Stevenson, R.A., Fister, J.K., Barnett, Z.P., Nidiffer, A.R., Wallace, M.T., 2012. Interactions between the spatial and temporal stimulus factors that influence multisensory integration in human performance. Exp. Brain Res. 219 (1), 121–137.

Stevenson, R.A., Wilson, M.M., Powers, A.R., Wallace, M.T., 2013. The effects of visual training on multisensory temporal processing. Exp. Brain Res. 225, 479–489.

Stickel, S., et al., 2019. Audio–visual and olfactory–visual integration in healthy participants and subjects with autism spectrum disorder. Hum. Brain Mapp. 40 (15), 4470–4486.

Talsma, D., Senkowski, D., Soto-Faraco, S., Woldorff, M.G., 2010. The multifaceted interplay between attention and multisensory integration. Trends Cogn. Sci. 14 (9), 400–410.

Tang, X., Wu, J., Shen, Y., 2016. The interactions of multisensory integration with endogenous and exogenous attention. Neurosci. Biobehav. Rev. 61, 208–224.

Taylor, K.I., Moss, H.E., Stamatakis, E.A., Tyler, L.K., 2006. Binding crossmodal object features in perirhinal cortex. Proc. Natl. Acad. Sci. 103 (21), 8239–8244.

Thurstone, L.L., 1919. The learning curve equation. Psychol. Monogr. 26 (3), i.

Tokoro, K., Sato, H., Yamamoto, M., Nagai, Y., 2015. Thalamus and attention. Brain and Nerve Shinkei Kenkyu No Shinpo 67 (12), 1471–1480.

Tomasi, D., Ernst, T., Caparelli, E.C., Chang, L., 2004. Practice-induced changes of brain function during visual attention: a parametric fMRI study at 4 Tesla. Neuroimage 23 (4), 1414–1421.

Tombari, D., et al., 2004. A longitudinal fMRI study: in recovering and then in clinically stable sub-cortical stroke patients. Neuroimage 23 (3), 827–839.

Tyll, S., Budinger, E., Noesselt, T., 2011. Thalamic influences on multisensory integration. Commun. Integr. Biol. 4 (4), 378–381.

Umarova, R.M., et al., 2010. Structural connectivity for visuospatial attention: significance of ventral pathways. Cereb. Cortex 20 (1), 121–129.

Valzolgher, C., Verdelet, G., Salemme, R., Lombardi, L., Gaveau, V., Farné, A., Pavani, F., 2020. Reaching to sounds in virtual reality: A multisensory-motor approach to promote adaptation to altered auditory cues. Neuropsychologia 149, 107665.

Van Atteveldt, N., Formisano, E., Goebel, R., Blomert, L., 2004. Integration of letters and speech sounds in the human brain. Neuron 43 (2), 271–282.

Van den Brink, R., Cohen, M.X., van der Burg, E., Talsma, D., Vissers, M., Slagter, H.A., 2014. Subcortical, modality-specific pathways contribute to multisensory processing in humans. Cereb. Cortex 24 (8), 2169–2177.

Van der Burg, E., Olivers, C.N.L., Bronkhorst, A.W., Theeuwes, J., 2008. Pip and pop: nonspatial auditory signals improve spatial visual search (in eng). J. Exp. Psychol. Hum. Percept. Perform. 34 (5), 1053–1065. https://doi.org/10.1037/0096-1523.34.5.1053, 2008/10//.

Van Wanrooij, M.M., Bell, A.H., Munoz, D.P., Van Opstal, A.J., 2009. The effect of spatial–temporal audiovisual disparities on saccades in a complex scene. Exp. Brain Res. 198 (2), 425–437.

Wang, F., Poston, T., Teo, C.L., Lim, K.M., Burdet, E., 2004. Multisensory learning cues using analytical collision detection between a needle and a tube. In: Proceedings of the 12th International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2004. IEEE, pp. 339–346. *HAPTICS'04*.

Ward, R., Arend, I., 2007. An object-based frame of reference within the human pulvinar. Brain 130 (9), 2462–2469.

Wuerger, S., Hofbauer, M., Meyer, G., 2003. The integration of auditory and visual motion signals at threshold. Percept. Psychophys. 65, 1188–1196.

Wuerger, S.M., Parkes, L., Lewis, P.A., Crocker-Buque, A., Rutschmann, R., Meyer, G.F., 2012. Premotor cortex is sensitive to auditory–visual congruence for biological motion. J. Cogn. Neurosci. 24 (3), 575–587.

Xu, W., Kolozsvari, O.B., Oostenveld, R., Hämäläinen, J.A., 2020. Rapid changes in brain activity during learning of grapheme-phoneme associations in adults. Neuroimage 220, 117058.

Yap, B.W., Sim, C.H., 2011. Comparisons of various types of normality tests. J. Stat. Comput. Simul. 81 (12), 2141–2155.

Yotsumoto, Y., Watanabe, T., Sasaki, Y., 2008. Different dynamics of performance and brain activation in the time course of perceptual learning. Neuron 57 (6), 827–833.

Zhang, J.Y., Zhang, G.L., Xiao, L.Q., Klein, S.A., Levi, D.M., Yu, C., 2010. Rule-based learning explains visual perceptual learning and its specificity and transfer. J. Neurosci. 30 (37), 12323–12328.

Zou, H., Müller, H.J., Shi, Z., 2012. Non-spatial sounds regulate eye movements and enhance visual search. J. Vis. 12 (5), 2. -2.