Virtual Reality (2025) 29:9
https://doi.org/10.1007/s10055-024-01086-w

ORIGINAL ARTICLE

Individual differences in processing multisensory information predict presence in different virtual reality environments

Callum P. Glenn[1] · Matthew Coxon[1]

Received: 12 March 2024 / Accepted: 9 December 2024

Abstract
One of the most notable aspects of using a head mounted display is the feeling of being 'within' the digitally created virtual world. Technological advances across the fields of computer science and engineering have helped to increase this sense of presence. However, there remain wide variations between individuals, which are only just beginning to be captured at a theoretical level. One potential explanation for these individual differences may be how individuals process multisensory information. In this paper we detail two studies which explored whether performance on two different multisensory tasks (the pip and pop and a redundant signals task), predict some of these differences in self-reported presence. In study 1 (N = 32), clear correlations between the multisensory task (pip and pop) and presence scores were found using an underwater simulation. This provided the first indication that individuals that were positively influenced by illusory audiovisual conjunctions were also those that had the strongest sense of presence. Study 2 (N = 68) expanded upon these results, finding again that multisensory integration (within a redundant signals task) was related to self-reports of presence using a different VR experience. In addition, it was found that this relationship could be moderated by either providing a unisensory or multisensory VR experience. Together these results suggest that individual differences in the efficiency of multisensory integration may provide an additional level of explanation not currently accounted for within theoretical models of presence formation.

Keywords Presence · Multisensory integration · Temporal binding windows · Pip and pop

1 Introduction

One of the most notable aspects of using a head mounted display (HMD) is the feeling of being 'in' the digitally created virtual world (e.g. Minsky 1980; Steuer 1992). Slater et al. (2022) emphasize that the feeling of 'being there' is one of two components that constitute a sense of presence within a virtual environment, known as the 'Place Illusion'. This feeling is more generally known as 'personal presence' (Heeter 1992), 'physical presence' (e.g. Lee 2004), or 'spatial presence' (e.g. Schubert et al. 2001; Wirth et al. 2007; Hartmann et al. 2015). Recent, advances in computer science and engineering have led to HMDs which can now evoke a particularly strong sense of spatial presence. Spatial presence is known to be influenced by the technical specifications of the HMD such as delay, field of view, quality of image and sound (see Cummings and Bailenson 2016, for review). However, when individuals use the exact same technology to explore a virtual environment, their experience of spatial presence still differs: spatial presence is influenced not only by the technical features of the technology, but also by the characteristics of the user.

Both technical factors and user factors are taken into account within the process model of spatial presence (Wirth et al. 2007). This process model of spatial presence consists of two levels. The first proposes that allocation of attention to spatial information, both automatic (influenced by the technical factors) and controlled (directed by the user); leads to the construction of a 'mental model'. The second level proposes that this 'mental model' is then subjected to perceptual hypothesis testing ("am I located within this environment?"). If the perceptual hypothesis is accepted with an affirmatory response, then the phenomenological experience of spatial presence is thought to occur. The

✉ Callum P. Glenn
c.glenn@yorksj.ac.uk

[1] School of Education, Languages and Psychology, York St John University, Lord Mayor's Walk, York YO31 7EX, UK

Published online: 20 December 2024

process model is noteworthy because, alongside technical factors, cognitive factors are also important in the formation of spatial presence, specifically, attention allocation and egocentric reference frames (see Riecke and Von der Heyde 2002). Interesting for the present purpose, Wirth et al. (2007) also identify the importance of multisensory integration (p. 500–503), although it does not become a central part of their model.

The idea that multisensory integration may have an influence on the formation of spatial presence is not new. At least seven different accounts of spatial presence have presented it as a stage or factor (Buset 2015; Kim and Biocca 1997; Lombard and Ditton 1997; Nunez 2007; Schubert 2009; Slater 2002; Wirth et al. 2007; Witmer and Singer 1998). Such a hypothesis is also consistent with findings within computer science, in which multisensory environments are more likely to evoke a sense of spatial presence in general. For example, Dinh et al. (1999) investigated the effects of four sensory cues (visual, auditory, tactile and olfactory) on reported feelings of spatial presence and found an additive effect when more were senses stimulated. Similar results are reported elsewhere within the literature (e.g. Davis et al. 1999; Insko 2001).

One potential explanation for why multisensory integration may influence spatial presence is due to one of the building blocks of spatial presence; egocentric reference frames (Wirth et al. 2007). Egocentric reference frames are mental models of the world from a first-person perspective (Mou and McNamara 2002). These models are developed from the sensory information an individual perceives from their surrounding environment. However, when sensory information from multiple modalities represents different environments (such as a VR world and a real-world lab), multiple ERFs may be developed (Riecke and von der Heyde 2002). Individuals are hypothesised to feel present in the most congruent ERF, referred to as the primary ERF (Wirth et al. 2007). Thus, if a virtual environment can provide more sensory information and an individual can integrate this multisensory information into a salient percept, they are more likely to develop a congruent ERF.

Whilst multisensory integration is thought to be important in the formation of spatial presence; findings within cognitive neuroscience have taught us that these abilities differ significantly between individuals. Independent of research with HMDs, it has been established that multisensory integration of highly synchronous information differs greatly between individuals (e.g. Donohue et al. 2010; Foss-Feig et al. 2010; Hillock-Dunn and Wallace 2012). Those who process it more efficiently have been documented as gaining behavioural advantages in such circumstances. For example, multisensory integration leads to faster and more accurate task responses in comparison to a single modality (Colonius and Diederich 2004; Diederich and Colonius 2004; Talsma and Woldorff 2005; Todd 1912; Van der Burg et al. 2008). On the other hand, individuals have also been noted to differ in how well they can maintain this multisensory integration, when the information is presented several milliseconds apart (Calvert et al. 2004). This is due to a window of processing time known as the temporal binding window (TBW), which exists to account for the differences in transmission and transduction speeds of different sensory information (Pöppel 1988; Stein and Meredith 1993). These differences in cognitive processing could potentially explain some of the variance in spatial presence experiences.

No published research could be identified that has investigated the effect of an individual's ability to integrate multisensory information on their feelings of spatial presence. However, recent studies have identified a relationship between multisensory integration and simulator sickness (e.g., Sadiq 2019; Sadiq and Barnett-Cowan 2022). Simulator sickness is a special form of motion sickness that can occur when using a HMD. Zilka and Bonneh (2022), identified a positive relationship between the width of an individual's temporal binding window and how susceptible participants were to motion sickness. Simulator sickness has also been related to the size of an individuals' temporal binding window, in the same direction as motion sickness (Sadiq 2019). In addition, Kim et al. (2020) demonstrated simulator sickness and spatial presence have an inverse relationship with scene instability and delay lag (concepts related to multisensory integration). There is therefore clear initial evidence that differences in multisensory integration may be related to simulator sickness. However, such relationships are yet to be determined with spatial presence directly.

In order to investigate this relationship a suitable measure of multisensory integration must be chosen. There are multiple ways to investigate individuals' proficiency of integrating multisensory information, including computational and psychophysical methods (Cornelio et al. 2021). However, behavioural experiments tend to be the most commonly used (Razavi et al. 2020). Here we collected data using two different behavioural experiments to understand an individuals' ability to integrate audio-visual information. Although individuals have the ability to integrate multiple different modalities (e.g. vestibular, haptic), all the systems are believed to be connected (e.g. Alais et al. 2010; Stein and Meredith 1993). Therefore, results pertaining to an individuals' ability to integrate audio-visual information, should provide a representation of their overall ability to integrate multisensory information.

The first measure used within this research involves identifying individuals' ability to integrate audio-visual conjunctions in the form of the pip and pop task (Van der Burg

et al. 2008). This task presents non-spatial auditory information at the same time as a target line changes colour. This target line is either horizontal or vertical and participants are required to determine it's orientation as fast as possible. If the information is integrated, it can either facilitate (decrease) or inhibit (increase) response times. Comparing the visual search times to conditions where there is no auditory information, provides an index of how well individuals integrate congruent multisensory information (Lui and Wong 2012).

In the first study we asked whether individuals who integrate multisensory information effectively are the same individuals who tend to feel more present whilst using a HMD. Participants first completed the pip and pop task, they then experienced a short virtual experience, in which they accompanied a navy seal on a virtual submarine ride within a HMD. The demonstration was stationary with non-spatial audio, in order to reduce the number of conflicting cues presented within the environment. Individuals then rated their sense of spatial presence within the simulation on two different measures. It was anticipated that performance on the cognitive task would be highly related to self-reports of spatial presence: individuals that benefit from effective multisensory integration may also feel more present in a multisensory virtual environment with limited conflicting information.

2 Study 1

2.1 Methods

The research followed the British Psychological Society guidelines for ethical practice, seeking written consent from participants and informing participants of their right to withdraw from the study at any point, including after data collection.

2.1.1 Participants

Thirty-two participants took part, with an average age of 20.56 years (SD = 1.79, 22 females). Participants were sampled opportunistically, ranging from university staff to undergraduate students. No incentives were offered for participation (e.g., no course credit or monetary reward).

2.1.2 Design

The main measures within the study were as follows: accuracy and average reaction time on the pip and pop task; self-reported experience of spatial presence using the spatial presence experience scale (SPES) questionnaire (Hartmann et al. 2016); and self-reported experience of presence using the Slater, Usoh and Steed (SUS) questionnaire (Slater et al. 1994). Demographic data concerning age and gender were also collected, as well as self-reports of previous gaming experience, and previous virtual reality experience.

2.1.3 Materials

2.1.3.1 Questionnaires In the study, two presence questionnaires were administered. Firstly, the spatial presence experience scale (SPES); which consists of 8 items adapted from two spatial presence related scales of the Measurement, Effects, Conditions (MEC) spatial presence questionnaire (MEC-SPQ; Vorderer et al. 2004) by Hartmann and colleagues (Hartmann et al. 2016). Out of the 8-item questionnaire, four items relate to self-location (e.g. "I felt like I was actually there in the environment of the presentation.") and four relate to perceived possible actions (e.g. "The objects in the presentation gave me the feeling that I could do things with them."). Each item is rated on a 5-point scale, where 1 resembles 'I do not agree at all' and 5 resembles 'I fully agree'. The SPES has a high internal consistency (Hartmann et al. 2016) and the interpretation of the data are directly related to the process model of spatial presence from which it was developed (Wirth et al. 2007).

The second questionnaire was the three item SUS-presence questionnaire using three questions which Slater et al. (1994) used to investigate the subjective experience of presence. Each item was rated on a 7-point Likert-type scale. Using the original continuous scoring strategy, the presence score is then generated by adding up the total of each response, producing a score out of 21. Whilst there have been several variations of the SUS (6 Item: Slater et al. 1995, 7 Item: Usoh et al. 2000), alongside different scoring strategies (continuous vs all-or-nothing), the three-item version was chosen because of its continuous scoring strategy and its high convergent validity with other measures (Youngblut 2003; Youngblut and Huie 2003). Although the SUS was originally developed to be used with older virtual reality systems, such as the 'DIVISION ProVision200 system', the questionnaire has been used to assess presence within similar HMD's as used in this study (e.g. Bormann 2005; Sayyad et al. 2020).

Alongside the two presence questionnaires, a measure of previous gaming experience was used to control for the potentially heightened multisensory ability of video game players (Donohue et al. 2010). Lemmens et al. (2011a, b) 'time spent on games' measure was chosen because it was easily adapted to include previous virtual reality experience. Therefore, a four-item questionnaire was developed, providing two continuous scores. The first item was adapted

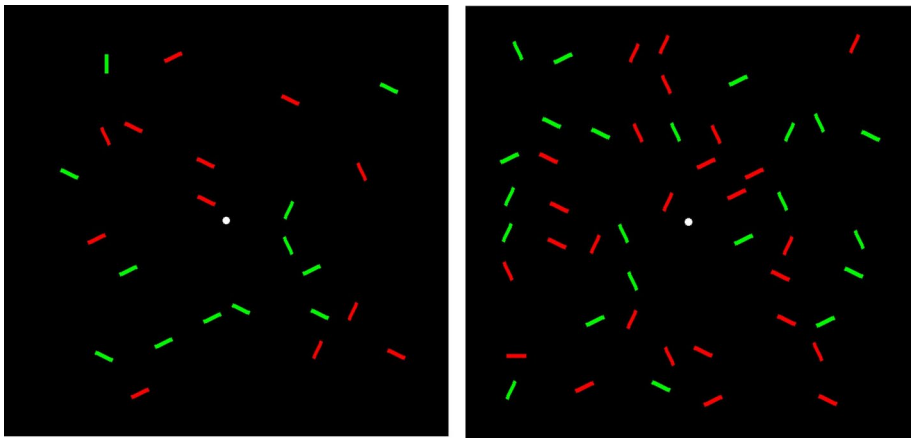

Fig. 1 Example fields of colour changing lines from Van der Burg et al's (2008) pip and pop task. These images are non-consecutive

Fig. 2 Example image from inside the SEAL delivery vehicle simulation. Images from SEAL Delivery Vehicle VR Experience, Received 01/02/17 from https://www.oculus.com/experiences/rift/1181842295168854/

to include mobile phones; "How many days a week do you play games on a console/pc/handheld/mobile phone". The second item was also adapted to include mobile phones; "On an average day that you play games on a pc/console/handheld/mobile phones, how much time do you spend playing?". The weekly time spent on games was measured by multiplying the days per week by the number of hours per day. The third item asked, "How many times have you used virtual reality equipment before this experiment?" followed by "How long did you spend using the virtual reality equipment on average per session? (Please leave blank if above answer is Zero)". The total time previously spent using virtual reality was calculated by multiplying the times used by average time spent using virtual reality.

2.1.3.2 Pip and pop task Illusory audio-visual conjunctions were measured using the pip and pop task, developed by Van der Burg et al. (2008). The task requires participants to search for a colour changing (red–green) vertical or horizontal line amongst a field of colour changing diagonal distractors. In half of the conditions, a non-spatial tone (500 Hz) with a 60 ms duration (5 ms fade-in and out) was provided at the same time the target line changed colour. All lines on the screen changed colour at least once every 900 ms, with the individual lines changing at a rate of either 50, 100 or 150 ms. The number of distractors fell into three conditions; either 24, 36 or 48 lines within the field (see Fig. 1). Each distractor line was offset by 22.5° clockwise or anticlockwise from vertical or horizontal. Both participant's accuracy of response (percentage of correctly identified target lines) and reaction time (ms) were measured for each condition. In keeping with the original research, there were 10 blocks of 48 trials and the trials alternated between sound conditions. Amid each block of trials, there was a break and the participants only moved to the next block when ready.

2.1.3.3 VR simulation The virtual reality experience was a 3° of freedom (DoF) underwater submarine journey with head tracking, but no user controls or virtual body. Participants were virtually placed next to a navy SEAL in scuba gear and could view two other people swimming near the vessel (see Fig. 2). The total length of the short fidelity demonstrator was approximately 3 min.[1] The experience was

[1] SEAL Delivery Vehicle VR Experience developed by Mass Virtual and distributed by Oculus.

chosen as it provided audio-visual information regarding an environment (no vestibular or haptic sensory information).

2.1.3.4 Equipment and VR hardware The hardware used to run the pip and pop task was a Dell 20″ monitor with a 1600 × 900 resolution, and sounds were played through Sony MDR-ZX110 headphones. The virtual reality experience was shown using the CV1 version of the Oculus rift. The head mounted display (HMD) had a resolution of 1200 × 1080 pixels per eye, an 110° field of view, and a 90 Hz refresh rate. The HMD provided both head and position tracking through infrared LEDs. It was connected to an Alienware X51 desktop with an Intel i5-6400 processor, an 8 GB dual channel DDR4 2133MHZ and NVIDIA GeForce GTX 970 with 4 GB GDDR5: VR-Ready graphics card.

2.1.4 Procedure

Participants were given instructions on what the study entailed, alongside a health and safety brief regarding the VR equipment. After completing a consent form, participants were required to enter their age and gender on the E-prime program (E-Prime 2.0; Psychology Software Tools, 2012) the pip and pop task was being conducted on. Participants wore a pair of headphones, and a tone sounded when the information screen loaded to indicate sound was working. The information screen informed the participants that they would have to find either a horizontal or a vertical line amongst a field of diagonal lines. Once identified, they were asked to press a corresponding key on the keyboard (Z for Vertical and M for horizontal). At this point, the researcher left the room, allowing the participant to complete the first task without distraction. Overall, there were 10 blocks of 48 trials, and the tasks lasted between 30 and 75 min. This was due to some individuals taking longer breaks in between each block of trials, in comparison to others.

After participants had completed the first task, they were given a 5 min break before the VR experience (for more details see Sect. 2.1.3.3 above). After the VR experience, the participants were offered a glass of water and a short break before completing the three questionnaires in order to mitigate feelings of simulator sickness. No participants required this break. The participants filled in the questionnaires in order, firstly the SUS, followed by the SPES and finally the questionnaire regarding previous gaming and virtual reality experience.

2.2 Results

E-prime (E-Prime 2.0; Psychology Software Tools, 2012) was used to calculate both the overall reaction time and accuracy means; and means for each set size, split between sound conditions for each participant. All statistical analysis was conducted using SPSS 29.0.

To assess whether the pip and pop effect was present within this group of participants, all data were subject to a repeated-measures ANOVA. The analysis was in line with the original authors (Van der Burg et al. 2008) with both set size (24, 36, 48) and sound condition (Sound, No-sound) being within-subject variables. Violations of sphericity were identified for the effect of set size; therefore, the reported p-values are after a Greenhouse–Geisser correction. However, neither the sound condition nor interaction effect violated sphericity.

In line with Van der Burg et al. (2008), average RT's were faster in the sound condition than in the no sound condition $F(1,31) = 12.15$, $p = 0.001$, $\eta = 0.28$. Average RT's increased significantly as the set size increased $F(1.46,45.06) = 95.24$, $p < 0.001$, $\eta = 0.75$. In addition, the interaction between sound and set size was also significant $F(2,62) = 6.65$, $p = 0.002$, $\eta = 0.18$, with sound showing a faster response across all 3 set sizes (Fig. 3).

Accuracy also displayed similar results to Van der Burg et al. (2008), with the sound condition and interaction displaying no significant effects. The accuracy of participants was found to significantly differ between set size after a Greenhouse–Geisser correction for sphericity violation $F(1.53,47.39) = 3.67$, $p = 0.044$, $\eta = 0.11$. However, further pairwise comparison with a Bonferroni correction identified no significant difference between any set size (Fig. 4).

Participant's multisensory integration index scores were therefore only generated for the measure of reaction time. Due to the interaction effect for the reaction time data being significant, the averages were collapsed over each of the set sizes to form an overall average reaction time for both sound conditions. Following the calculation outlined by Lui and Wong (2012); (No Sound RT–Sound RT)/No Sound RT), MSI indices for each participant ranged from − 0.54 to 0.55 ($\bar{X} = 0.15$, $SD = 0.24$), with positive values indicating multisensory facilitation and negative values indicating multisensory inhibition. For example, those with a MSI index of 0.50 responded twice as quick on average in the sound condition.

Raw questionnaire data was converted to total scores in the following ways: the questions composing the SUS were added together to give a continuous score out of 21, with the minimum score being 3. The SPES formed 3 scores; A self-location scale which was calculated by adding the first four questions providing a score between 4 and 20. A possible perceived actions scale which was a sum of the final

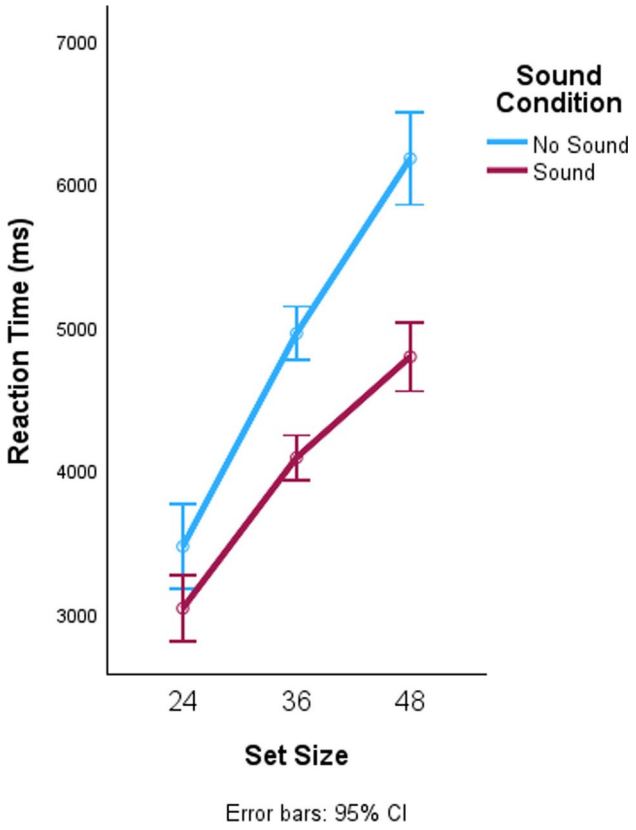

Fig. 3 Results for mean reaction time from the Pip and Pop task by set size and sound condition

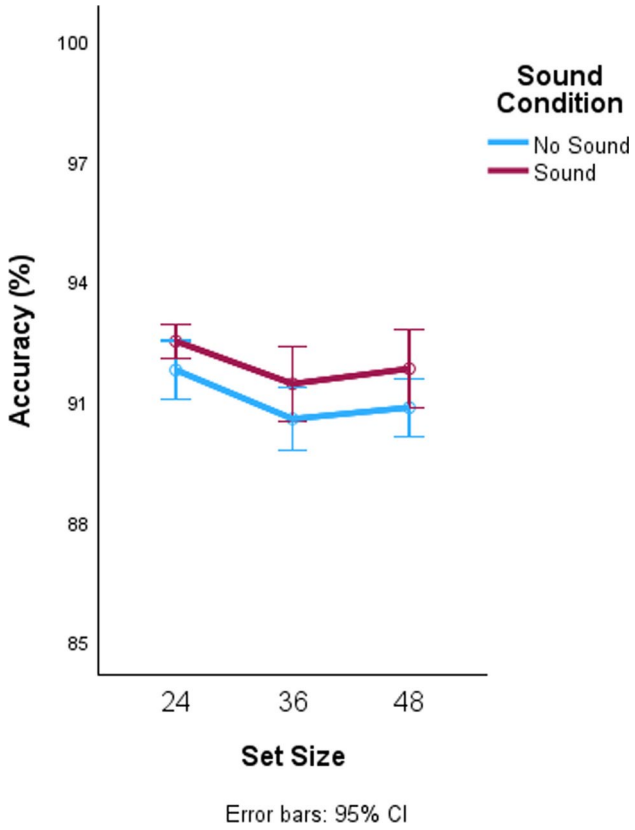

Fig. 4 Results for search accuracy from the Pip and Pop task by set size and sound condition

four questions, again providing a score between 4 and 20. Finally, a total SPES score was formed through adding both self-location and possible perceived action scores, which formed a minimum of 8 and a maximum of 40. The previous gaming experience of each participant was calculated by multiplying the average daily usage with the average amount of days spent playing a week, providing a continuous score. Previous VR experience was calculated in a similar manner, through multiplying the average amount of time spent within virtual reality per session, by the number of sessions. Descriptive statistics for the measures are displayed below in Table 1.

Nonparametric correlations were calculated between the main presence measures and the control measures (gaming experience, VR experience, age and gender) to identify if there were any relationships that needed to be statistically controlled. Significant relationships were identified between age and SPES PPA ($p(32) = -0.360$ $p = 0.021$), and age and SPES total ($p(32) = -0.311$ $p = 0.041$). Age was therefore identified as an important variable to control for in the analysis. Gender, gaming experience, and VR experience did not have a significant relationship with any presence measures, or the MSI index, and were therefore not included in any further analysis.

Table 1 Descriptive statistics for questionnaire measures (N = 32)

	Measure	Mean	Standard deviation	Minimum	Maximum
Presence	SUS	14.16	3.73	4.00	19.00
	SPES: SL	14.06	3.66	4.00	20.00
	SPES: PPA	13.44	3.29	7.00	19.00
	SPES Total	27.50	6.31	11.00	38.00
Previous experience	Gaming	4.80	6.07	0.00	21.00
	Virtual reality	0.74	2.09	0.00	10.50

SUS, Slater, Usoh and Steed Presence Questionnaire; SPES, spatial presence experience scale; SL, self location; PPA, perceived possible actions

Before running partial correlations, the dataset was checked for outliers on the presence questionnaires. Two participant's data fell outside 2.5 standard deviations from the mean on three of the presence measures and were therefore excluded. Partial correlations between the reaction time MSI index and the presence measures were run, controlling for age (Fig. 5). There were significant moderate partial correlations between participants overall MSI RT index and the total SPES score ($r(27) = 0.374$, $n = 30$, $p = 0.046$), the total SUS score ($r(27) = 0.412$, $n = 30$, $p = 0.027$), and the Perceived Possible Action scale ($r(27) = 0.438$, $n = 30$, $p = 0.018$) whilst controlling for age. However, the Self

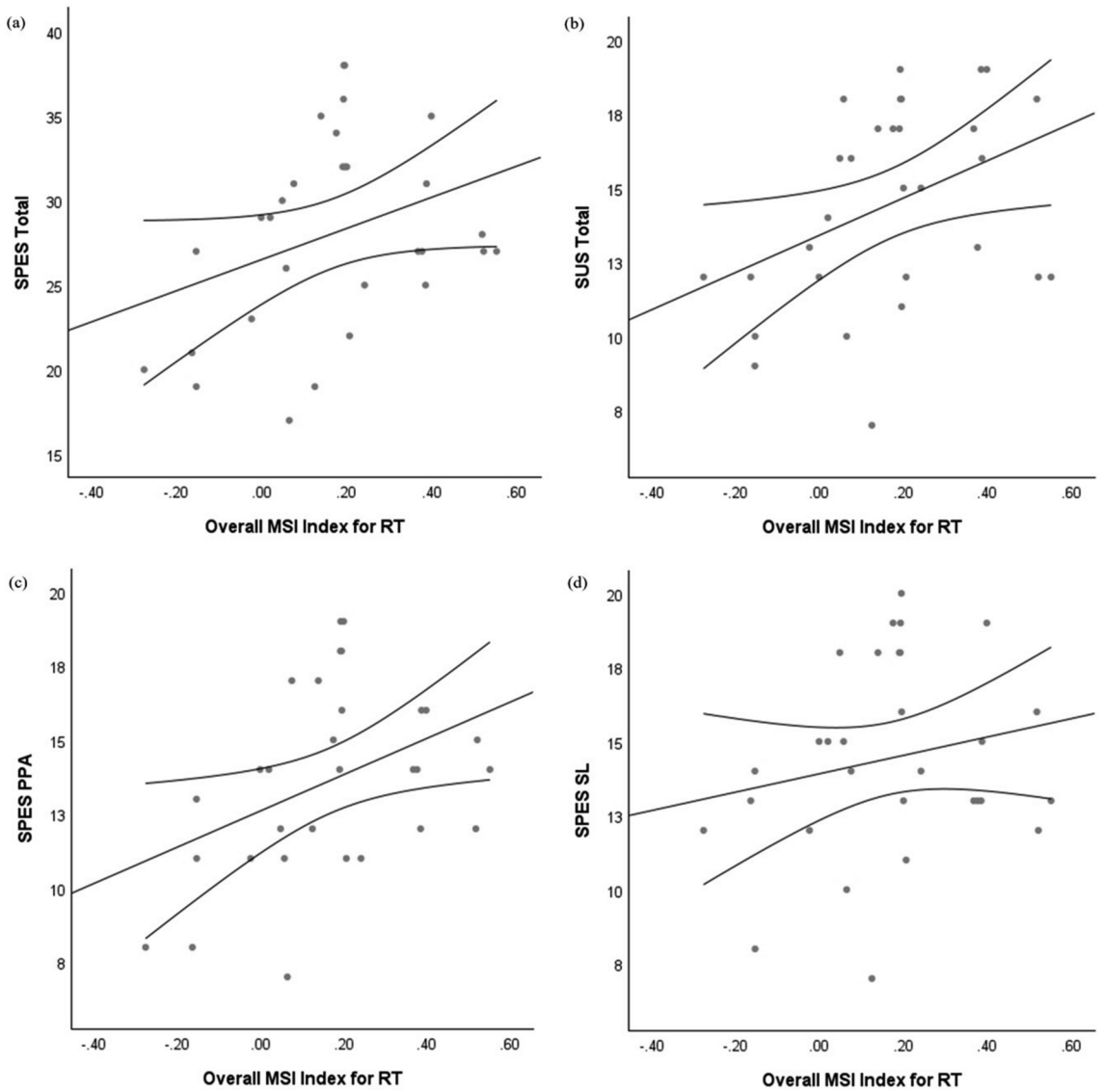

Fig. 5 Scatterplots of the partial correlations between Overall MSI RT index and SPES Total (**a**), SUS Total (**b**), SPES PPA (**c**) and SPES SL (**d**)

Location scale did not significantly correlate with participants overall MSI RT index ($r(27) = 0.229$, $n = 30$, $p = 0.233$).

In summary, behavioural data from the MSI measures (reaction time and accuracy) were consistent with previous research confirming their validity. Accuracy was identified to be affected by set size, however pairwise comparisons with a Bonferroni correction did not identify any significant differences between the conditions. Therefore, an overall MSI index was only calculated for the reaction time data. Nonparametric correlations identified age as a confounding variable, therefore age was included in the subsequent partial correlations. These partial correlations identified a statistically significant and positive relationship between the majority of the presence measures and the MSI RT index.

2.3 Discussion

In this study we investigated the relationship between individuals' ability to integrate multisensory information and feelings of spatial presence in a stationary virtual environment. Participants that performed faster in the multisensory condition of the pip and pop task also self-reported higher levels of spatial presence in a subsequent simulation. This indicates, for the first time, that differences in individual experience of spatial presence can be linked to individual differences in ability to integrate multisensory information.

Whilst the virtual reality environment (VRE) used within this study involved both audio and visual sensory cues, which mirrored the modalities of the multisensory integration task, it is possible that the results are specific to the environment used. Particularly given that the auditory information

presented in the VRE was neither spatially nor temporally aligned to a visual stimulus which is less likely to occur in other simulations. The exploratory nature of study one also meant that the sample size was relatively small, although statistically significant relationships were still found. Study two addressed these issues of generalisability. A larger number of participants experienced a room-scale VRE and were assigned to one of two groups, with one experiencing the VRE in a multisensory condition and the other experiencing a visual-only unisensory condition. This allowed for the comparison of different levels of sensory information within the same VRE. The separation into groups aimed to avoid any cross-over effects which might arise from completing both conditions. If differences in multisensory integration are related to spatial presence, then the presence or absence of multisensory information in the simulation should moderate this relationship.

In addition to adjusting the environment participants experienced, a second measure of multisensory integration was included. The first study relied upon audio-visual conjunctions which demonstrate an individual's multisensory gains when presented with congruent multisensory information. The following study also used a redundant signals task to assess how offset information can be presented and still demonstrate these multisensory gains within a cognitive process known as a time window of integration/temporal binding window (Colonius and Diederich 2004, 2020). As these temporal binding windows allow individuals to cope with offset information, it is hypothesised that those with wider windows will feel more present in virtual environments where multisensory information is presented (as this will naturally be offset).

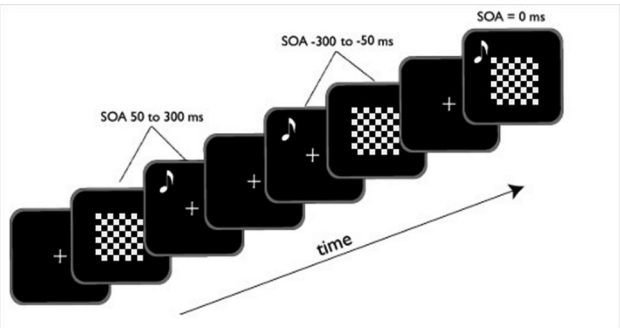

Fig. 6 Example task procedure for the redundant signal task, demonstrating range of stimuli offset asynchronies

3 Study 2

3.1 Methods

The research followed the British Psychological Society guidelines for ethical practice, seeking written consent from participants and informing participants of their right to withdraw from the study at any point, including after data collection.

3.1.1 Participants

Sixty-eight participants (47 females, mean age 21.38, SD 7.40) were recruited through a course participation scheme at York St John University. All participants were awarded course credits for taking part. Each participant was randomly assigned to one of two groups. '*Sound-Present*' which experienced the VRE with both audio and visual information (34 participants). The other group, '*Sound-Absent*' experienced the VRE with only visual information (34 participants).

3.1.2 Design

The main measures within this study were identical to the measures presented in study 1, except participants temporal binding windows were also measured using a redundant signals task. In addition, we only included set sizes 24 and 48 from the pip and pop task to reduce potential fatigue effects. Participants who were assigned to the '*Sound-Present'* group experienced the VRE with both audio and visual information. The other group '*Sound-Absent*', experienced the VRE with only visual information. In both groups, haptic feedback was disabled in an attempt to reduce additional sensory input.

3.1.3 Materials

3.1.3.1 Redundant signals task Participants temporal binding windows were measured using a flash-beep redundant signals task coded in Inquisit Millisecond (Inquisit 6, 2020). The visual stimuli consisted of a 5×5(cm) black and white checkerboard which appeared in front of a fixation cross on a black background (see Fig. 6). This was presented approximately 60 cm in front of the participants on a Lenovo Legion Y25f-10 monitor, which had 62.23 cm screen size and a 144 Hz refresh rate. The auditory stimuli were a 440HZ pure tone with a 33 ms duration presented binaurally via Sony MDR-ZX110 headphones. Participants were asked to respond as quickly as possible to whichever stimuli were presented first, by pressing the spacebar. The stimuli were offset in intervals of 50 ms ranging from 300 ms audio leading (bleep presented before the checkerboard), up to 300 ms

visually leading (checkerboard presented before the bleep) (Fig. 6). Each stimuli offset asynchrony was repeated randomly 12 times, leading to a total of 156 trials over the 13 intervals.

3.1.3.2 VR simulation and hardware The simulation participants experienced was 'Job Simulator: The 2050 Archives',[2] (see Fig. 7). For this study, participants completed the 'Office Worker' simulation. Throughout the simulation participants are instructed to complete a range of tasks; from choosing new employees to sending emails. All the while being confined to an office cubicle which matched the dimensions of the lab room the experiment was taking place within (2.3 m by 4 m). The maximum length of time spent within the VRE was 30 min. The VR equipment used in this study was updated to the Oculus Quest, which is a standalone head mounted display (HMD) with a resolution of 1440×1600 pixels per eye, a 93° field of view, and a 72 Hz refresh rate. The HMD provided both head and position tracking through an "inside-out" tracking system. The HMD uses a Qualcomm Snapdragon 835 'System-on-Chip' with 4GB of RAM and an Adreno 540 GPU. Within the virtual environment participants could control both movement and interactions using the Oculus Quest wireless controllers. These are tracked using a simultaneous localization and mapping system using AI, accelerometers, and the built-in camera to track infrared diodes on the controllers. For participants experiencing the 'sound-*absent*' environment, instructions were provided through subtitles. In both groups, haptic feedback from the controls was disabled.

3.1.3.3 Equipment The hardware used to run the pip and pop and redundant signals task was a 24.5inch Lenovo Legion Y25f-10 monitor with a 1920×1080 resolution, and sounds were played through Sony MDR-ZX110 headphones.

3.1.3.4 Procedure Participants completed this study in an analogous way to study 1 (see Sect. 2.1.4). First information regarding the study was provided to the individuals, before consent and demographic details were recorded. After which participants donned the headphones, before completing the pip and pop task and the redundant signals task in a counterbalanced order.

Once participants had completed the tasks, they were offered a 5-min break, before being instructed on how to equip and operate the Oculus Quest. This included a quick demonstration from the researcher in how to operate the controls, adjust the head straps for comfort, and the lens width for clarity. Participants entered the HMD with the experience preloaded in a 'staging area', which required individuals to pick up a specified item and place it in a slot to begin the experience. The 'staging area' acted as a way to confirm the headset was correctly set up for the individual and that they understood how to operate the simulation. The Oculus Quest's guardian boundary (safe area) was set to 2 m×3.5 m and the researcher was quietly present during the entire simulation for safeguarding. Each participant

Fig. 7 Example image from inside the Office Worker Simulation. Images from Job Simulator: The 2050 Archives, Received 09/08/23 from https://jobsimulatorgame.com/

[2] A video game developed by Owlchemy Labs and distributed by Meta.

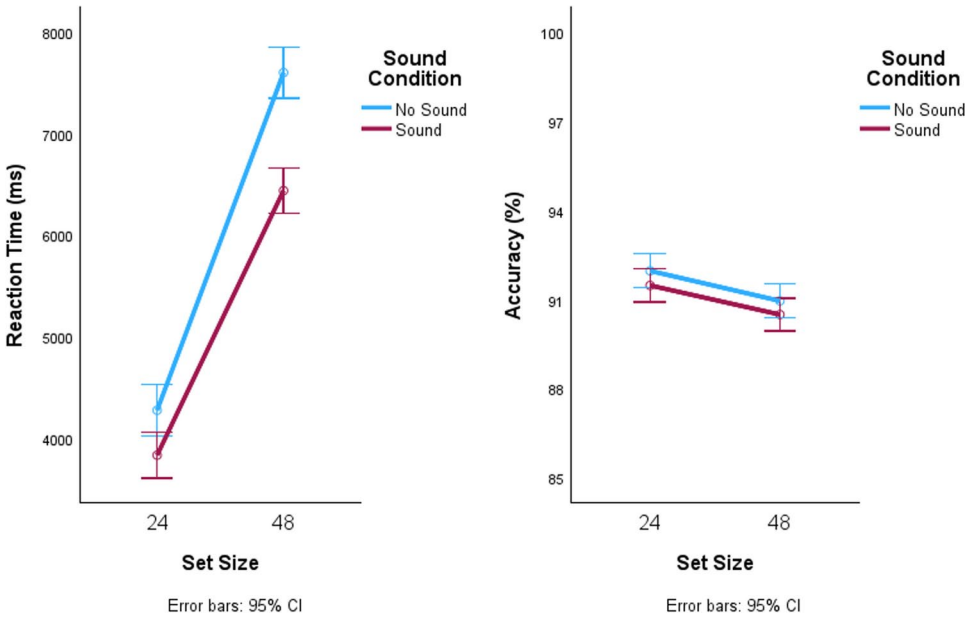

Fig. 8 Results for mean reaction time and accuracy on the pip and pop task comparing a 24 set size to 48 by sound condition in Study 2

spent 30 min within the simulation before receiving a drink of water and completing the same three questionnaires as previously administered in study 1.

3.2 Results

All data from the pip and pop task were subject to a repeated-measures ANOVA first to validate the behavioural task. Set size (24, 48) and sound condition (Sound or No sound) were entered as the within-subject variables. Average RT's were faster in the sound condition than in the no sound condition $F(1,67) = 14.47$, $p < 0.001$, $\eta = 0.18$. Average RT's increased significantly as the set size increased $F(1,67) = 225.33$, $p < 0.001$, $\eta = 0.77$. In addition, the interaction between sound and set size was also significant $F(1,67) = 6.91$, $p = 0.011$, $\eta = 0.09$, with sound showing a faster response across both set sizes (Fig. 8). Accuracy was also explored with a repeated measure ANOVA. Participants accuracy was found to significantly differ between set size $F(1,67) = 7.64$, $p = 0.007$, $\eta = 0.10$. However, both the sound condition and interaction displayed no significant effects. Therefore, participants MSI indices were only calculated for participants overall reaction times in line with study one. The overall MSI RT index ranged from −0.47 to 0.68 (X = 0.10, SD = 0.26).

Raw data from the redundant signal task were checked for outliers (< 30 ms and > 3000 ms) with individual responses being removed. The window widths for each participant were then calculated using MATLAB scripts adapted from Diederich and Colonius' (2015) twin model. The scripts use *fmincon* to optimise the objective function:

Table 2 Descriptive statistics for questionnaire measures used in Study 2 (N = 65)

	Measure	Mean	Standard deviation	Minimum	Maximum
Presence	SUS	15.28	3.34	6.00	21.00
	SPES: SL	16.00	3.00	9.00	20.00
	SPES: PPA	17.00	2.57	11.00	20.00
	SPES Total	33.00	4.83	20.00	40.00
Previous experience	Gaming	11.36	17.61	0.00	98.00
	Virtual reality	6.77	28.14	0.00	200.00

SUS, Slater, Usoh and Steed Presence Questionnaire; SPES, spatial presence experience scale; SL, self location; PPA, perceived possible actions

$$\sum_{\tau} \left(\frac{mean\left[RT_{VA,\tau}\right] - E\left[\widehat{RT_{VA,\tau}}\right]}{SE\left(mean\left[RT_{VA,\tau}\right]\right)} \right)^2$$

where mean[$RT_{VA,\tau}$] relates to the mean reaction times for each SOA (τ) in the reaction time task and $E[\widehat{RT_{VA,\tau}}]$ are the predicted values for each SOA based on the twin models parameters: visual processing time (λV), auditory processing time (λA), the common processing stage (μ), the TBW for the task (ω), and multisensory temporal gain (Δ). Therefore, the objective function calculates the deviation between measured and estimated RT means, returning the fit with the smallest deviation and identifying the TBW based on the twin models calculations. Participants predicted TBW's ranged from 15.65 to 740.65 ms ($\bar{X}$ = 47.13 ms, SD = 90.77 ms).

Raw questionnaire data for the SPES, SUS and gaming experience questionnaire were converted into total scores in the same way as in the previous study. Each measure was

checked for extreme outliers (over 2.5SD from the mean). Three participants were removed from further analysis due to being classed as an extreme outlier on multiple variables. Descriptive statistics for the measures are reported below (Table 2).

Nonparametric correlations were conducted between the MSI measures, the Presence scores, and the control variables (Age, Gender, VR and Gaming experience) to identify whether any relationships between the variables needed to be controlled for in the following analysis. No significant correlations were identified between the control measures and the main variables.

In order to identify whether multisensory environments moderated the relationship between an individual's RT MSI index or TBW and their sense of presence, moderated multiple regressions were conducted for each presence measure. Assumptions of moderated multiple regression were checked. The residuals were uncorrelated and distributed normally and there was no multicollinearity. However, the data displayed homoscedasticity. Therefore heteroscedasticity-consistent standard errors were used within the multiple moderated regression using Hayes Process module in SPSS (model 1). The results for each scale are discussed below.

3.2.1 Spatial presence experience scale

The first regression investigated the relationship between the MSI index and SPES total score. There was a nonsignificant main effect found between sound condition and the total SPES score, b = − 0.141, CI [− 1.15, 0.873], t = − 0.278, $p = 0.782$, and nonsignificant main effect of RT MSI index on the total SPES score (b = 0.551, CI [− 3.237, 4.338], t = 0.291, $p = 0.772$). However, there was a significant interaction found by sound condition on RT MSI index and total SPES, b = − 6.800, CI [− 10.587, − 3.013], t = − 3.5903, $p < 0.001$. It was found that participants who experienced a unisensory VRE displayed a positive relationship between the RT MSI index and the total SPES score (b = 7.351, CI [1.925, 12.776], t = 2.709, $p = 0.009$), whereas those who experienced a multisensory VRE displayed a negative relationship between the RT MSI index and total SPES score (b = − 6.249, CI [− 11.535, − 0.9636], t = − 2.364, $p = 0.021$). From these results, it can be concluded that the relationship between illusory audio-visual conjunctions and spatial presence is moderated by the multisensory nature of the VRE (Fig. 9).

The second regression explored the relationship between individuals temporal binding windows and the SPES total score. There was a significant relationship between individuals TBW and the total SPES score (b = 0.033, CI [0.172, 0.049], t = 4.152, $p < 0.001$), although there was not a significant relationship between the sound condition and the SPES score (b = 0.498, CI [− 0.579, 1.576], t = 0.925, $p = 0.360$). However, these effects were moderated by a significant interaction between sound condition and RT TBW (b = 0.045, CI [0.029, 0.061], t = 5.592, $p < 0.001$). The results highlighted that as the size of individuals TBWs increased in the sound condition so did their SPES Score (b = 0.078, CI [0.046, 0.109], t = 4.919, $p < 0.001$), whereas the increase in TBWs in the no sound condition had a smaller negative effect on individuals SPES score (b = − 0.012, CI [− 0.016, − 0.007], t = − 5.235, $p < 0.001$). This also indicates that the relationship between spatial presence and individuals binding of multisensory information is moderated by the multisensory nature of the VR experience (Fig. 10).

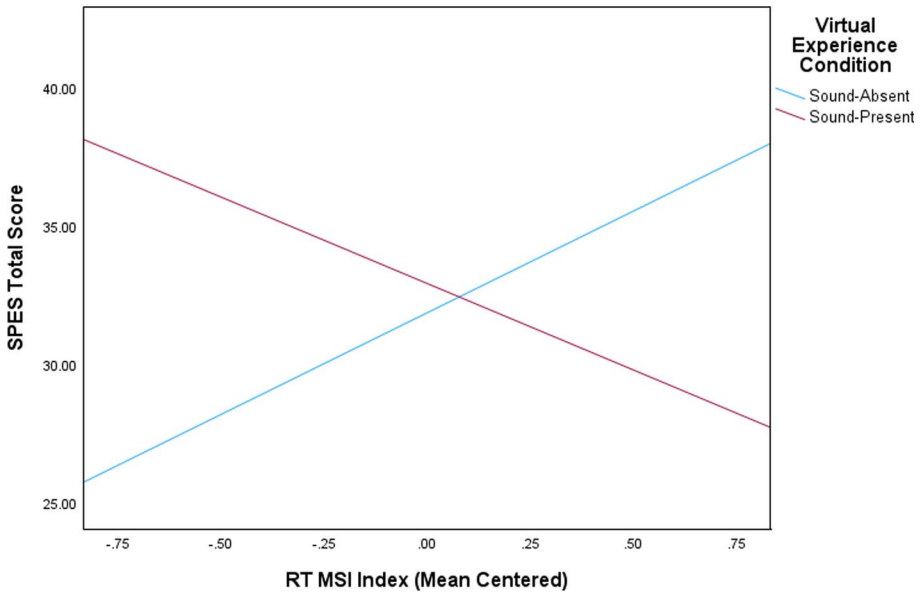

Fig. 9 Interaction effect of the presence of sound on the relationship between reaction time multisensory integration index and self-reported spatial presence score on the spatial presence experience scale

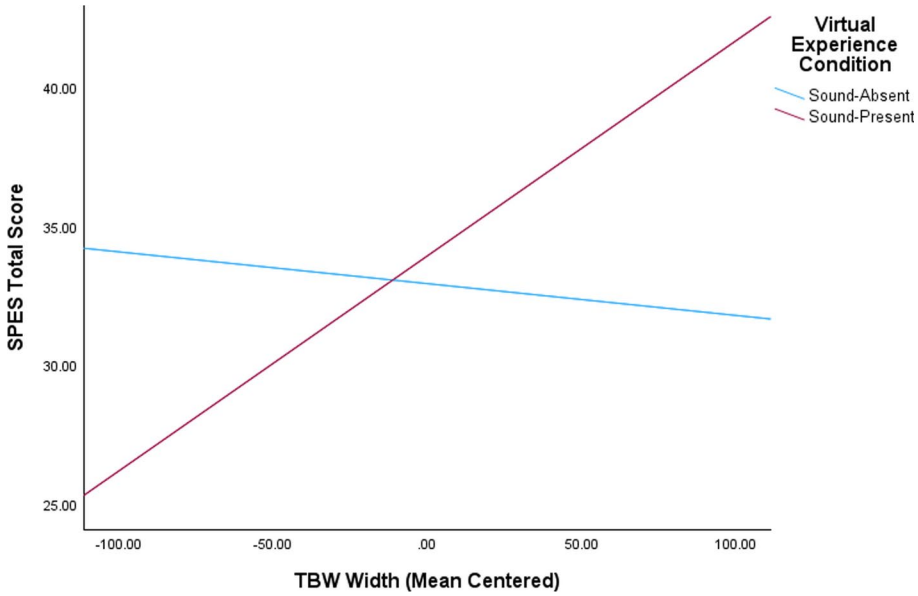

Fig. 10 Interaction effect of the presence of sound on the relationship between individuals temporal binding window and self-reported spatial presence score on the spatial presence experience scale

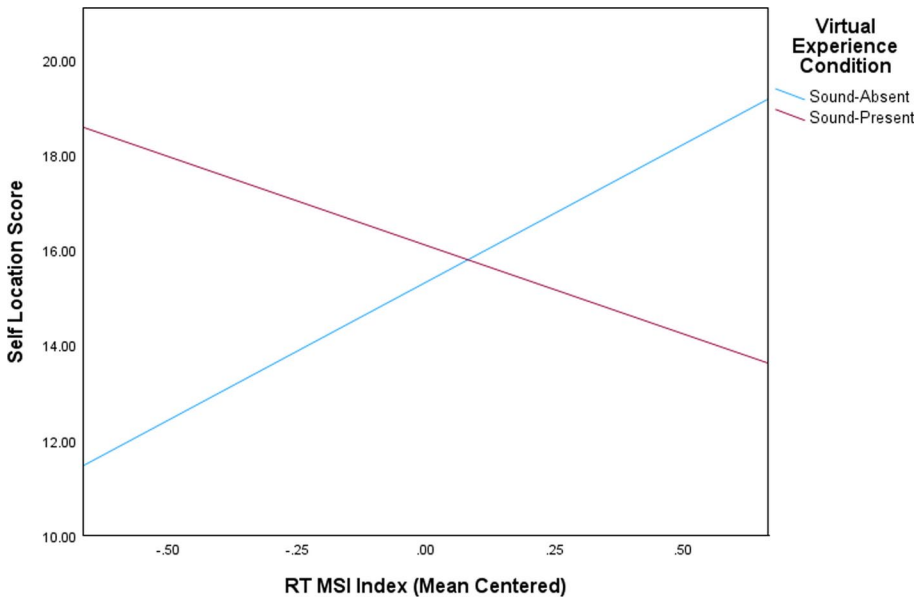

Fig. 11 Interaction effect of the presence of sound on the relationship between reaction time multisensory integration index and self-reported score on the self-location subscale

3.2.2 Slater, Usoh and Steed presence questionnaire

Unlike the SPES measure of presence, there were no significant main or interaction effects between the sound condition and participants RT MSI index and the total SUS score. The same is true for the effect of TBW on total SUS score, with no main or interaction effects being identified as statistically significant ($p < 0.05$).

3.2.3 SPES subscales: self location and perceived possible actions

To further investigate the moderated relationship between participants ability to integrate multisensory information and the SPES score, four additional MMRs were conducted on both of the measure's subscales (Self Location; and Perceived Possible Actions). There was a nonsignificant main effect found between sound condition and Self Location score, b = 0.390, CI [− 0.232, 1.011], t = 1.253, $p = 0.215$, and a nonsignificant main effect of RT MSI index on the total SPES score (b = 1.034, CI [− 1.382, 3.450], t = 0.855, $p = 0.396$). However, there was a significant interaction found by sound condition on RT MSI index and Self Location, (b = − 4.761, CI [− 7.177, − 2.345], t = − 3.940, $p < 0.001$). It was found that participants who experienced a unisensory VRE displayed a positive relationship between the RT MSI index and Self Location score (b = 5.794, CI [1.878, 9.710], t = 2.959, $p = 0.004$), whereas those who experienced a multisensory VRE displayed a negative relationship between the RT MSI index and Self Location score

(b = −3.727, CI [−6.558, −0.896], t = −2.633, $p = 0.011$) (Fig. 11).

Alongside the MSI index, there was a significant relationship found between TBW and Self Location score, b = 0.011, CI [0.002, 0.021], t = 2.462, $p = 0.017$, and a significant relationship between the sound condition and the self location score, b = 0.784, CI [0.107, 1.461], t = 2.316, $p = 0.024$. The analysis also identified a significant interaction of sound condition on TBW and self location, b = 0.020, CI [0.011, 0.029], t = 4.361, $p < 0.001$. It was found that participants who experienced a unisensory VRE displayed a negative relationship between the TBW and self location score (b = −0.009, CI [−0.011, −0.006], t = −7.443, $p < 0.001$), whereas those who experienced a multisensory VRE displayed a positive relationship between the TBW and Self Location score (b = 0.032, CI [0.013, 0.050], t = 3.439, $p = 0.001$) (Fig. 12).

In addition, there was also a significant relationship found between TBW and Perceived Possible Actions score, b = 0.022, CI [0.013, 0.031], t = 4.930, $p < 0.001$, but a non-significant relationship between the sound condition and the Perceived Possible Actions score, b = −0.286, CI [−0.848, 0.276], t = −1.017, $p = 0.31$. However, the analysis did identify a significant interaction of sound condition on TBW and Perceived Possible Actions, b = 0.025, CI [0.016, 0.033], t = 5.545, $p < 0.001$. It was found that participants who experienced a unisensory VRE displayed a small negative relationship between the TBW and perceived possible actions score (b = −0.003, CI [−0.005, −0.0002], t = −2.134, $p = 0.037$), whereas those who experienced a multisensory VRE displayed a positive relationship between the TBW and perceived possible actions score (b = 0.046, CI [0.029, 0.064], t = 5.294, $p < 0.001$) (Fig. 13). However, there were no significant main or interaction effects of sound condition and participants RT MSI index on the perceived possible actions scale. It can therefore be concluded that the moderating effect of sound on the relationship between illusory audio-visual conjunctions and spatial presence only affected participants sense of 'being there' (not 'doing there') in this particular VRE.

In summary, several moderated multiple regressions were conducted to identify if the relationship between an individuals' multisensory integratory ability and their self-reported sense of spatial presence, was moderated by the virtual environment containing audio-visual or visual-only information. Whilst there were no significant effects when using the SUS to measure presence, the relationship between the SPES and participants MSI index and TBW's was significantly moderated by sound condition. This supported the hypothesis that it is how individuals process multisensory information that has an important relationship with their development of a sense of spatial presence.

3.3 Discussion

When individuals experienced a room-scale virtual environment which only contained visual information, individuals who did better with congruent information in the pip and pop task felt more present. Alongside those who had a narrower temporal binding window. However, individuals who had a wider temporal binding window felt more present in a multisensory environment. In addition, those who did better when presented with simultaneous audio-visual information in the pip and pop task, felt less present. The two tasks differ in how they measure an individual's ability to handle multisensory information. The pip and pop task provides a measure of individuals' multisensory gain (facilitation or inhibition) when presented with congruent stimuli. The

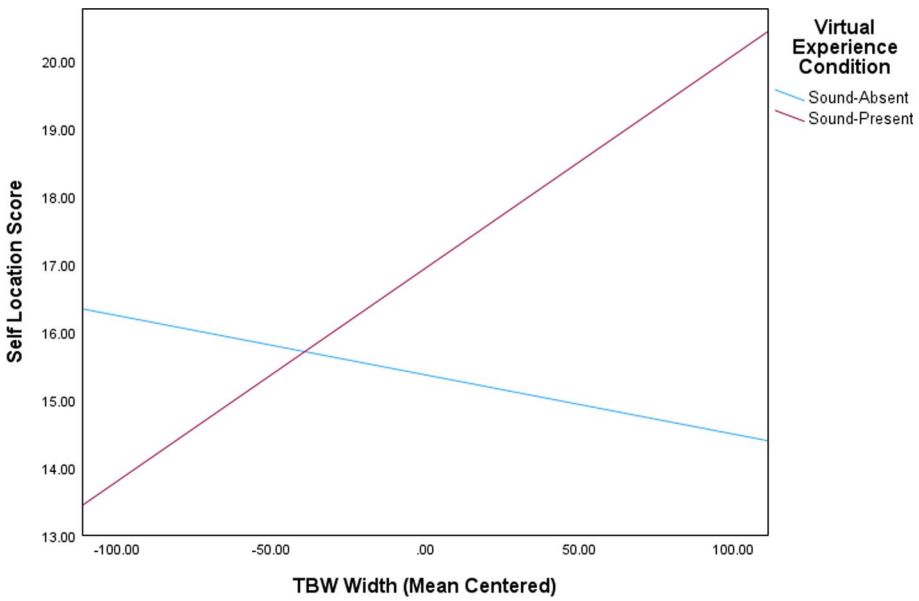

Fig. 12 Interaction effect of the presence of sound on the relationship between individuals temporal binding window and self-reported score on the self-location subscale

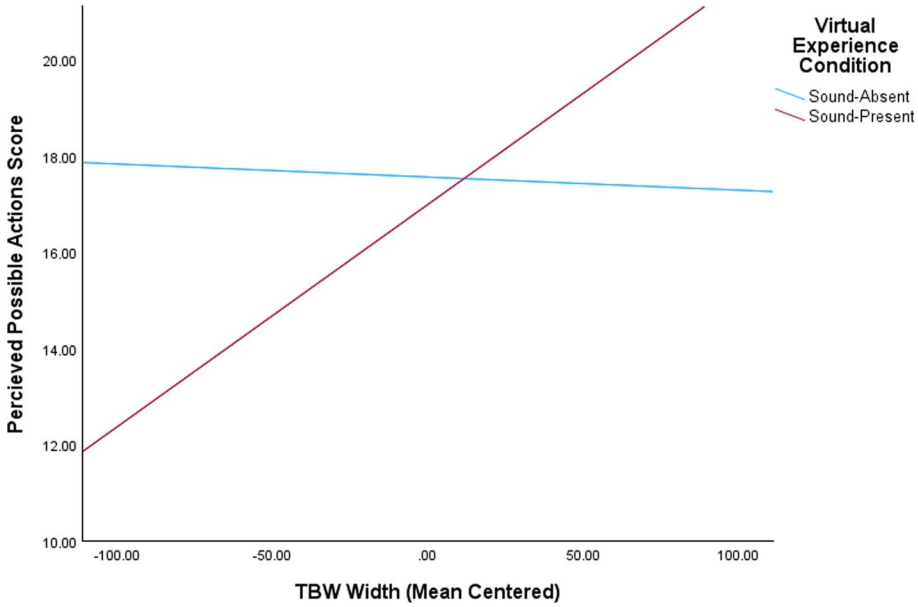

Fig. 13 Interaction effect of the presence of sound on the relationship between individuals temporal binding window and self-reported score on the perceived possible action subscale

redundant signals task measures how offset stimuli can be presented and still demonstrate multisensory gains associated with binding the stimuli into a single percept. This suggests two separate processes are occurring; one related to the speed of processing congruent stimuli (Stationary/Visual only), and another related to maintaining a unified percept when information is offset (Roomscale/Multisensory).

Whilst there have been technological advances in display lag and offsets in VR headsets in the past 20 years, the results from the research presented here may be explained by the fact that there will always be discrepancy between modality update times (e.g., Dixit and Sarangi 2023). Even when controlling for a small offset (4 ms) in VR headsets, scene instability still has an impact on spatial presence (Kim et al. 2020). As much as HMD developers will endeavour to reduce this latency to as minimal an offset as possible, it is unlikely to ever reach true congruency between all the senses, with less than 16.7 ms currently being considered good (Bailey et al. 2004; Jung et al. 2000). In simulations which allow individuals to move within the environment (Roomscale), this latency is likely to be larger than in a passive experience where the individual is sat stationary. Especially when the experience is multisensory. Therefore, it is important to consider that spatial presence differs between individuals in line with their ability to integrate multisensory information, and how this might impact design decisions.

The results from these studies raise several important suggestions for the understanding of how individuals' multisensory integratory abilities affect their sense of spatial presence. Firstly, individuals who experience gains in visual field search response times when presented with congruent auditory information go on to feel more present in virtual environments with low-conflicting information (stationary/unisensory). Secondly, individuals who have a wider window for binding multisensory information (less likely to notices delays) feel more present in a multisensory virtual environment, than those with a narrower temporal binding window. The results from both studies therefore indicate that the facilitation of processing multisensory information quicker than unisensory information in real life tasks may lead to quicker development of the necessary mental models and perceptual hypothesis testing, in environments where conflicting sensory information is kept to a minimum. When this is not the case, those with wider tolerances for offset multisensory information can maintain the mental model of the virtual environment more sufficiently.

Although different environmental factors may have effects on spatial presence (e.g. Stanney et al. 2002), multisensory integration only occurs whenever two or more modalities are stimulated within an individuals' temporal binding window. Thus, if an environment produces congruent multisensory information, the impact on individuals' ability to integrate this information would remain the same across different environments with the same number of congruent multisensory features. The relationship between multisensory integration and spatial presence, as identified here, should therefore not be constrained by specific environmental factors, but by a combination of the extent to which these factors fail to be presented congruently, and an individual's tolerance for how incongruent these factors are. For example, in study 1, the individual was sat stationary (the vehicle moved) and limited to 3° of freedom. This produces less incongruent information than in the multisensory experience in study 2, where the individual was standing and could move with 6° of freedom. The key finding is therefore that when experiencing environments with different levels

of potentially conflicting information; spatial presence differs between individuals, consistent with differences in their MSI ability.

Furthermore, whilst multisensory integration is not directly considered in Wirth and colleagues (Wirth et al. 2007) process model, it lends support to the assumptions surrounding the consistency of multisensory cues in virtual environments. This evidence suggests that individual differences of multisensory integration should be considered in models of spatial presence, more so than the requirement that environments provide consistent multisensory cues (e.g., Bailey et al. 2004; Steuer 1992), as it lends one explanation as to why an individuals level of spatial presence differs, within the same virtual experience. Therefore, we propose an extension to the process model which includes more specific details about the user factors affecting both the formation of a spatial situational model, and the role of the primary egocentric reference frame within the model (Hofer et al. 2012; Wirth et al. 2007) which we believe to be the primary mechanism multisensory integration is impacting. Whilst the model adequately covers the importance of the mediated environment presenting 'consistent' multimodal information in order to support both the construction of a coherent mental model and the perceptual hypothesis testing of this model as the primary 'egocentric reference frame' (Riecke and Von der Heyde 2002), the model does not elaborate on an individuals' ability to process this consistent information beyond"…*internal processes based on user characteristics also affect the construction and quality of the SSM"* (Wirth et al. 2007, p.504). Hence, we suggest that this should necessarily include individual differences in ability to handle offset multisensory information through an increased binding window, alongside the differences in facilitation of response times to congruent sensory information. This would provide a deeper understanding of how the spatial situational model may be formed. This is in line with past research which has highlighted the need for breadth, depth, and consistent sensory information (e.g. Steuer 1992; Witmer and Singer 1998). In addition, it has the potential to explain how the spatial situational model might be maintained in some individuals but not others when multimodal information is not consistent, thus causing the generation of multiple ERF's (Wirth et al. 2007).

It's important to acknowledge some limitations of the studies presented here. Firstly, the opposite relationship between presence and individuals' MSI index in studies 1 and 2 is believed to be due to the differences in the sensory information generated within the VREs. However, due to the nature of the variations between the VREs (such as stationary vs room-scale, passive vs active, 3-DoF vs 6-DoF), other influencing factors may be in play. While it has been suggested that two separate processes are likely occurring, further research could be conducted to better isolate these mechanisms. Secondly, despite both studies yielding statistically significant results, the small participant samples in both studies mean that the results reported here are exploratory in nature. These studies are among the first to explore the relationship between individuals' MSI ability and presence, so it's important for future replications to be carried out to build upon these findings. Nonetheless, despite these limitations, the results presented here propose an extension to the process model of spatial presence (Wirth et al. 2007), one that considers individual differences in ability to process multisensory information.

Future research should aim to address these limitations and proposals. Firstly, by expanding upon the results identified here, with a specific focus on individuals' ability to bind offset information. This might include investigating whether stationary or room-scale virtual environments produce different results due to the nature of stationary "open world" locomotion and conflicting vestibular sensory information. In addition, the psychophysical literature regarding temporal binding windows has also highlighted that individuals' may be able to adjust their temporal binding window either automatically based on task demands (Diederich and Colonius 2015; Mégevand et al. 2013) or after perceptual training (e.g. Powers et al. 2009). Therefore, an alternative avenue of study is to investigate whether individuals binding windows 'recalibrate' to cope with the different representation of sensory information within a HMD, and whether this can be manipulated prior to experiencing a virtual environment to artificially increase individual's sense of spatial presence through manipulation of these underlying cognitive processes.

3.4 Conclusion

The aim of the research presented here was to investigate whether differences in multisensory integration, between different users, were consistent with differences in spatial presence when using an HMD. Using the pip and pop and a redundant signal task as our measures, we found that individuals who have improved visual search response times to congruent multisensory information, feel more present in low-conflicting virtual environments, but not in potentially conflicting environments. In addition, it was established that individuals who had a wider temporal binding window and therefore could tolerate greater offsets between sensory information, felt more present in multisensory environments. From these results, we propose an extension to the process model of spatial presence, to include individuals' ability to process multisensory information as a key user characteristic of the development of spatial presence thus, demonstrating that when experiencing the same virtual

environment; spatial presence differs between individuals, consistent with differences in their MSI ability.

Author contribution C.G—conceptualisation, methodology, software, formal analysis, investigation, data curation, writing—original draft, visualisation. M.C—conceptualisation, writing—reviewing and editing, supervision.

Data availability Data that support the findings of this research have been deposited in the Research at York St John Data Repository, accessed via the following links: https://figshare.com/s/16e80c4ae87bd904d4ff, https://figshare.com/s/3616ae82f36598848e77.

Declarations

Conflict of interest The authors declare no competing interests.

References

Alais D, Newell F, Mamassian P (2010) Multisensory processing in review: from physiology to behaviour. Seeing Perceiving 23(1):3–38. https://doi.org/10.1163/187847510X488603

Bailey RE, Arthur III JJ, Williams SP (2004) Latency requirements for head-worn display S/EVS applications. In: Proceeding SPIE, 5424, Enhanced and Synthetic Vision 2004, (11 August 2004). https://doi.org/10.1117/12.554462

Bormann K (2005) Presence and the utility of audio spatialization. Presence Teleoper Virtual Environ 14(3):278–297. https://doi.org/10.1162/105474605323384645

Buset M (2015) Can living in virtual environments alter reality?. In: 2015 IEEE virtual reality conference, VR 2015 - proceedings, pp 413–418. https://doi.org/10.1109/VR.2015.7223467

Calvert GA, Spence C, Stein BE (Eds.) (2004) The handbook of multisensory processes. Boston Review. https://doi.org/10.7551/mitpress/3422.001.0001

Colonius H, Diederich A (2004) Multisensory interaction in saccadic reaction time: a time-window-of- integration model. J Cogn Neurosci 16(6):1000–1009. https://doi.org/10.1162/0898929041502733

Colonius H, Diederich A (2020) Formal models and quantitative measures of multisensory integration: a selective overview. Eur J Neurosci 51(5):1161–1178. https://doi.org/10.1111/ejn.13813

Cornelio P, Velasco C, Obrist M (2021) Multisensory integration as per technological advances: a review. Front Neurosci. https://doi.org/10.3389/fnins.2021.652611

Cummings JJ, Bailenson JN (2016) How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. Media Psychol 19(2):272–309. https://doi.org/10.1080/15213269.2015.1015740

Davis ET, Scott K, Pair J, Hodges LF, Oliverio J (1999) Can audio enhance visual perception and performance in a virtual environment? Proc Hum Factors Ergon Soc Annu Meet 43(22):1197–1201. https://doi.org/10.1177/154193129904302206

Diederich A, Colonius H (2004) Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. Percept Psychophys 66(8):1388–1404. https://doi.org/10.3758/BF03195006

Diederich A, Colonius H (2015) The time window of multisensory integration: relating reaction times and judgments of temporal order. Psychol Rev 122(2):232–241. https://doi.org/10.1037/a0038696

Dinh HQ, Walker N, Hodges LF, Song C, Kobayashi A (1999) Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. In: Proceedings - virtual reality annual international symposium, pp 222–228. https://doi.org/10.1109/vr.1999.756955

Dixit A, Sarangi SR (2023) Minimizing the motion-to-photon-delay (MPD) in virtual reality systems. *ArXiv E-Prints*. https://doi.org/10.48550/arXiv.2301.10408

Donohue SE, Woldorff MG, Mitroff SR (2010) Video game players show more precise multisensory temporal processing abilities. Atten Percept Psychophys 72:1120–1129. https://doi.org/10.3758/APP.72.4.1120

Foss-Feig JH, Kwakye LD, Cascio CJ, Burnette CP, Kadivar H, Stone WL, Wallace MT (2010) An extended multisensory temporal binding window in autism spectrum disorders. Exp Brain Res 203(2):381–389. https://doi.org/10.1007/s00221-010-2240-4

Hartmann T, Wirth W, Vorderer P, Klimmt C, Schramm H, Böcking S (2015) Spatial presence theory: state of the art and challenges ahead. In: Immersed in media: telepresence theory, measurement and technology, pp 1–332. https://doi.org/10.1007/978-3-319-10190-3

Hartmann T, Wirth W, Schramm H, Klimmt C, Vorderer P, Gysbers A, Böcking S, Ravaja N, Laarni J, Saari T, Gouveia F, Sacau AM (2016) The spatial presence experience scale (SPES): a short self-report measure for diverse media settings. J Media Psychol 28(1):1–15. https://doi.org/10.1027/1864-1105/a000137

Heeter C (1992) *Being* there: the subjective experience of presence. Presence Teleoper Virtual Environ 1(2):262–271. https://doi.org/10.1162/pres.1992.1.2.262

Hillock-Dunn A, Wallace MT (2012) Developmental changes in the multisensory temporal binding window persist into adolescence. Dev Sci 15(5):688–696. https://doi.org/10.1111/j.1467-7687.2012.01171.x

Hofer M, Wirth W, Kuehne R, Schramm H, Sacau A (2012) Structural equation modeling of spatial presence: the influence of cognitive processes and traits. Media Psychol 15(4):373–395. https://doi.org/10.1080/15213269.2012.723118

Insko BE (2001) Passive haptics significantly enhances virtual environments. Doctoral Dissertation, The University of North Carolina at Chapel Hill. ProQuest Dissertations & Theses Global

Jung JY, Adelstein BD, Ellis SR (2000) Discriminability of prediction artifacts in a time-delayed virtual environment. Proc Hum Factors Ergon Soc Annu Meet 44(5):499–502. https://doi.org/10.1177/154193120004400504

Kim T, Biocca F (1997) Telepresence via television: two dimensions of telepresence may have different connections to memory and persuasion. J Comput-Med Commun. https://doi.org/10.1111/j.1083-6101.1997.tb00073.x

Kim J, Luu W, Palmisano S (2020) Multisensory integration and the experience of scene instability, presence and cybersickness in virtual environments. Comput Hum Behav 113(January):106484. https://doi.org/10.1016/j.chb.2020.106484

Lee KM (2004) Presence, explicated. Commun Theory 14(1):27–50. https://doi.org/10.1111/j.1468-2885.2004.tb00302.x

Lemmens JS, Valkenburg PM, Peter J (2011a) Psychosocial causes and consequences of pathological gaming. Comput Hum Behav 27(1):144–152. https://doi.org/10.1016/j.chb.2010.07.015

Lemmens JS, Valkenburg PM, Peter J (2011b) The effects of pathological gaming on aggressive behavior. J Youth Adolesc 40:38–47. https://doi.org/10.1007/s10964-010-9558-x

Lombard M, Ditton T (1997) At the heart of it all: the concept of presence. J Comput-Med Commun. https://doi.org/10.1111/j.1083-6101.1997.tb00072.x

Lui KFH, Wong ACN (2012) Does media multitasking always hurt? A positive correlation between multitasking and multisensory integration. Psychon Bull Rev 19(4):647–653. https://doi.org/10.3758/s13423-012-0245-7

Mégevand P, Molholm S, Nayak A, Foxe JJ (2013) Recalibration of the multisensory temporal window of integration results from changing task demands. PLoS ONE. https://doi.org/10.1371/journal.pone.0071608

Minsky M (1980) Telepresence. Omni, pp 45–51. https://doi.org/10.1351/pac198052010233

Mou W, McNamara TP (2002) Intrinsic frames of reference in spatial memory. J Exp Psychol Learn Mem Cogn 28(1):162. https://doi.org/10.1037/0278-7393.28.1.162

Nunez D (2007) A capacity limited, cognitive constructionist model of virtual presence. Doctoral Dissertation, University of Cape Town AUST Institutional Repository. http://repository.aust.edu.ng/xmlui/handle/11427/6426

Pöppel E (1988) Mindworks: time and conscious experience. (Artin T, Trans). Harcourt Brace Jovanovich

Powers AR, Hillock AR, Wallace MT (2009) Perceptual training narrows the temporal window of multisensory binding. J Neurosci 29(39):12265–12274. https://doi.org/10.1523/JNEUROSCI.3501-09.2009

Razavi M, Yamauchi T, Janfaza V, Leontyev A, Longmire-Monford S, Orr J (2020) Multimodal-multisensory experiments: design and implementation. *BioRxiv*. https://doi.org/10.1101/2020.12.01.405795

Riecke B, Von der Heyde M (2002) Qualitative modeling of spatial orientation processes using logical propositions: interconnecting spatial presence, spatial updating, piloting, and spatial cognition. Max Planck Institute for Biological Cybernetics, 100. https://www.researchgate.net/publication/216055708

Sadiq O (2019) Psychometric correlates of multisensory integration as potential predictors of cybersickness in virtual reality. Master's Thesis, University of Waterloo Waterloos International Repository. http://hdl.handle.net/10012/14896

Sadiq O, Barnett-Cowan M (2022) Can the perceived timing of multisensory events predict cybersickness? Multisens Res 35(7–8):623. https://doi.org/10.1163/22134808-bja10083

Sayyad E, Sra M, Höllerer T (2020) Walking and teleportation in wide-area virtual reality experiences. IEEE Int Symp Mixed Augment Real (ISMAR) 2020:608–617. https://doi.org/10.1109/ISMAR50242.2020.00088

Schubert TW (2009) A new conception of spatial presence: once again, with feeling. Commun Theory 19(2):161–187. https://doi.org/10.1111/j.1468-2885.2009.01340.x

Schubert T, Friedmann F, Regenbrecht H (2001) The experience of presence: factor analytic insights. Presence Teleoper Virtual Environ 10(3):266–281. https://doi.org/10.1162/105474601300343603

Slater M (2002) Presence and the sixth sense. Presence Teleoper Virtual Environ 11(4):435–439. https://doi.org/10.1162/105474602760204327

Slater M, Usoh M, Steed A (1994) Depth of presence in virtual environment. Presence Teleoper Virtual Environ 3(2):130–144. https://doi.org/10.1162/pres.1994.3.2.130

Slater M, Usoh M, Chrysanthou Y (1995) The influence of dynamic shadows on presence in immersive virtual environments. In: Göbel M (ed) Virtual environments '95. Eurographics. Springer, Vienna. https://doi.org/10.1007/978-3-7091-9433-1_2

Slater M, Banakou D, Beacco A, Gallego J, Macia-Varela F, Oliva R (2022) A separate reality: an update on place illusion and plausibility in virtual reality. Front Virtual Real. https://doi.org/10.3389/frvir.2022.914392

Stanney KM, Kingdon KS, Graeber D, Kennedy RS (2002) Human performance in immersive virtual environments: effects of exposure duration, user control, and scene complexity. Hum Perform 15(4):339–366. https://doi.org/10.1207/S15327043HUP1504_03

Stein BE, Meredith MA (1993) The merging of the senses. The MIT Press

Steuer J (1992) Defining virtual reality: dimensions determining telepresence, communication in the age of virtual reality. J Commun 42(4):73–93. https://doi.org/10.1111/j.1460-2466.1992.tb00812.x

Talsma D, Woldorff MG (2005) Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. J Cogn Neurosci 17(7):1098–1114. https://doi.org/10.1162/0898929054475172

Todd JW (1912) Reaction to multiple stimuli. The Science Press. https://doi.org/10.1037/13053-000

Usoh M, Catena E, Arman S, Slater M (2000) Using presence questionnaires in reality. Presence Teleoperators Virtual Environ 9(5):497–503. https://doi.org/10.1162/105474600566989

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

Vorderer P, Wirth W, Gouveia FR, Biocca F, Saari T, Jäncke L, Böcking S, Schramm H, Gysbers A, Hartmann T, Klimmt C, Ravaja N, Sacau A, Baumgartner T, Jäncke P (2004) MEC spatial presence questionnaire (MEC-SPQ). Short documentation and instructions for application. Report to the European Community, Project Presence: MEC (IST-2001–37661). https://doi.org/10.13140/RG.2.2.26232.42249

Wirth W, Hartmann T, Böcking S, Vorderer P, Klimmt C, Schramm H, Saari T, Laarni J, Ravaja N, Gouveia FR, Biocca F, Sacau A, Jäncke L, Baumgartner T, Jäncke P (2007) A process model of the formation of spatial presence experiences. Media Psychol 9(3):493–525. https://doi.org/10.1080/15213260701283079

Witmer BG, Singer MJ (1998) Measuring presence in virtual environments - a presence questionaire. Presence Teleoper Virtual Environ 7(3):225–240. https://doi.org/10.1162/105474698565686

Youngblut C, Huie O (2003) The relationship between presence and performance in virtual environments: results of a VERTS study. In: IEEE virtual reality, 2003. Proceedings, pp 277–278. https://doi.org/10.1109/VR.2003.1191158

Youngblut C (2003) Experience of presence in virtual environments. (Publication No. ADA427495) Institute for Defense Analyses. https://apps.dtic.mil/sti/citations/ADA427495

Zilka R, Bonneh Y (2022) Individual differences in audio-visual binding can predict the varied severity of motion sickness. *BioRxiv* https://doi.org/10.1101/2022.05.09.491170