Multisensory Integration With a Head-Mounted Display: Background Visual Motion and Sound Motion

William J. Harrison, Matthew B. Thompson, and **Penelope M. Sanderson,**
The University of Queensland, St Lucia, Queensland, Australia

Objective: The aim of this study was to assess how background visual motion and the relative movement of sound affect a head-mounted display (HMD) wearer's performance at a task requiring integration of auditory and visual information. **Background:** HMD users are often mobile. A commercially available speaker in a fixed location delivers auditory information affordably to the HMD user. However, previous research has shown that mobile HMD users perform poorly at tasks that require integration of visual and auditory information when sound comes from a free-field speaker. The specific cause of the poor task performance is unknown. **Method:** Participants counted audiovisual events that required integration of sounds delivered via a free-field speaker and vision on an HMD. Participants completed the task while either walking around a room, sitting in the room, or sitting inside a mobile room that allowed separate manipulation of background visual motion and speaker motion. **Results:** Participants' accuracy at counting target audiovisual events was worse when participants were walking than when sitting at a desk, $p = .032$. Compared with when they were sitting at a desk, participants' accuracy at counting target audiovisual events showed a trend to be worse when they experienced a combination of background visual motion and the relative movement of sound, $p = .058$. **Conclusion:** Multisensory integration performance is least effective when HMD users experience a combination of background visual motion and relative movement of sound. Eye reflexes may play an important role. **Application:** Results apply to situations in which HMD wearers are mobile when receiving multimodal information, as in health care and military contexts.

INTRODUCTION

Optical see-through head-mounted displays (HMDs) superimpose a continuous representation of visual information in the forward field of view, which can be convenient when work requires mobility (Laramee & Ware, 2002; Liu, Jenkins, & Sanderson, 2009; Sanderson et al., 2008). For example, HMDs have been shown to facilitate patient monitoring in operating theaters: HMD-wearing anesthesiologists spend more time attending to the patient if patient information is displayed on an HMD compared with traditional monitors alone (Liu, Jenkins, Sanderson, Watson, et al., 2009).

If auditory information is needed in addition to HMD-based information, however, it is unclear how the auditory information can be presented to the HMD wearer in a compelling yet cost-effective way. Extending the aforementioned example, anesthesiologists are required to monitor a patient's blood oxygen level via pulse oximetry, which consists of a visual waveform and an auditory tone from the monitor. If an anesthesiologist chooses to view the pulse oximetry waveform via HMD, then integration of HMD-based information with auditory information is important. In addition, very little research has been done on the effect of self-motion on people's ability to bind visual and auditory information under such circumstances.

Projecting a sound to the exact physical location of a virtual visual object on an HMD is difficult. Such sound projections require calculating the perceived location of the HMD

Address correspondence to Professor Penelope Sanderson, School of Psychology, The University of Queensland, St. Lucia QLD, Australia 4072; pennys@itee.uq.edu.au. ***HUMAN FACTORS,*** Vol. 52, No. 1, February 2010, pp. 78–91. DOI: 10.1177/0018720810367790. Copyright © 2010, Human Factors and Ergonomics Society.

object, calculating head-related transfer functions (Zotkin, Duraiswami, Grassi, & Gumerov, 2006) and then using sophisticated technology to project the sound to the exact location of the virtual object (Veltman, Oving, & Bronkhorst, 2004). Simple headphones, earpieces, or free-field sound are inexpensive alternatives, but we are only just starting to know about their advantages and disadvantages when used with HMDs (Thompson & Sanderson, 2008).

In the current study, we assessed factors that might influence how well people can integrate vision from an HMD and sound from a free-field speaker. When the wearer of an optical see-through HMD moves from place to place, the background visual environment moves relative to the images on the HMD, and sound delivered via a fixed speaker moves relative to the wearer and relative to the HMD images. If we control for any workload associated with walking, how might a moving background and relative motion of sound affect a person's ability to integrate vision from an HMD and sound from a speaker? In a recent review, Väljamäe (2009) indicates that we lack a clear understanding of how such visual and auditory phenomena interact even in basic laboratory conditions.

PREVIOUS RESEARCH

There is little prior research on multisensory integration for mobile workers. A recent experiment by Thompson and Sanderson (2008) tested whether the method of sound delivery affects people's ability to integrate multimodal information when they are walking and using an HMD. Thompson and Sanderson created the "mismatch task," in which participants viewed three shapes moving within a square wall on the HMD (see Figure 1). The shapes looked hard or soft, and they collided with each other and with the sides of the square wall. With each collision, there was a corresponding hard or soft sound, but occasionally there was a mismatch between the type of collision seen and the noise heard. Participants' accuracy at counting the number of mismatches was taken to indicate how easily people integrated sound and vision. Participants found it harder to count the mismatches when walking than when sitting if sound was delivered via a free-field speaker but not if sound was delivered via earpiece. Neither the fact of walking nor the use of a free-field speaker could individually account for the results. No full theoretical explanation was provided.

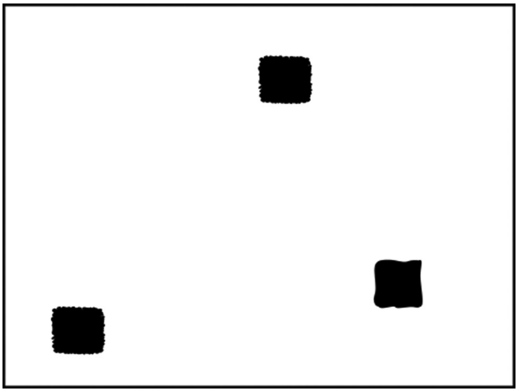

Figure 1. Moving shapes as seen on the HMD. The top shape and lower left shape are soft shapes, whereas the bottom right shape is a hard shape. The border is considered a hard object.

Source. From Figure 1 of "Multisensory Integration With a Head-Mounted Display: Sound Delivery and Self-Motion," by M. B. Thompson and P. M. Sanderson, 2008, *Human Factors*, *50*, p. 791. Copyright 2008 by Human Factors and Ergonomics Society. Used with permission.

If mobility makes it harder for people to integrate auditory and visual stimuli under certain circumstances, we need to understand when and how. Previous research on HMDs has often overlooked the possible influence of wearer mobility on task performance (Laramee & Ware, 2002). In the sections that follow, we discuss possible explanations for the Thompson and Sanderson (2008) result, which we explored in the experiment reported here.

Sound Motion

Multimodal stimuli are more likely to be integrated when they are presented in the same location rather than different locations in space and when they have common fate (for a review, see Bertelson & de Gelder, 2004). When an HMD user moves around, the spatial relationships change between vision on the HMD and the sound coming from a fixed speaker, so there are position and direction differences between auditory and visual stimuli. For example, the azimuth of the sound changes dynamically, whereas the HMD is always in the forward field of view.

Previous research indicates that when auditory and visual stimuli occur simultaneously, but in unpredictable and spatially separated locations, people sometimes perceive the stimuli to have appeared in sequence (Slutsky & Recanzone, 2001; Zampini, Guest, Shore, & Spence, 2005). During Thompson and Sanderson's (2008) mismatch task, if the sound and vision of a single event no longer seem simultaneous, then people may fail to detect a mismatch, or the sound may be incorrectly attributed to an adjacent visual event if the events occur in quick succession. Relative sound motion, therefore, may affect people's ability to integrate sound and vision by creating inconsistent and somewhat unpredictable spatial separation of auditory and visual stimuli.

Researchers have shown that saccadic and smooth-pursuit eye movements, both of which are used to monitor the shapes on the HMD in the mismatch task, can be impaired if sound is presented from a location changing relative to a visual event. As the spatial distance between a visual target and a concurrent auditory stimulus increases, participants take longer to initiate a saccade toward the visual target (Colonius & Arndt, 2001). Moreover, an increase in demand on spatial attention caused by sound motion can lead to impairments of smooth pursuits (Hutton & Tegally, 2005). Overall, impairments in eye movements caused by a moving sound source could decrease performance accuracy in an HMD-based task that requires close monitoring of visual events.

Background Visual Motion

In general, background visual motion stimulates eye reflexes, which could make it difficult for people to maintain fixation when using an optical see-through HMD. For example, optokinetic nystagmus (OKN) occurs in response to a visual scene moving in one direction, as experienced by someone standing on a train station platform when a train moves past (Leigh & Zee, 1999; Sharpe & Johnston, 1993). Attentionally demanding tasks can make suppression of OKN difficult (Williams, Mulhall, Mattingley, Lueck, & Abel, 2006). If an HMD user fails to suppress such a reflex response to background visual motion, fixation on the HMD may be lost and the user may fail to detect events on the HMD until he or she fixates again on the HMD.

Only recently have researchers begun to investigate the interaction between a moving visual environment and moving sounds. Väljamäe (2009) demonstrates that moving sounds may enhance vection, an illusory feeling of self-motion induced by a constantly moving visual environment. Teramoto, Watanabe, and Umemura (2008) and Teramoto, Watanabe, Umemura, Matsuoka, and Kita (2004) show that in conditions of vection, people misjudge the temporal order of visual events, of sounds, and of tactile stimuli, but they have not examined the effect of vection on multisensory integration across any of these modalities. Väljamäe speculates that the directional congruence between a moving sound and moving visual environment may be crucial for the perception of motion but that any interaction between vision, audition, and other sensory modalities that could influence basic perception is poorly understood.

Overall, background visual motion does not fully explain Thompson and Sanderson's (2008) findings. When listening to speaker sound, participants performed the mismatch task worse when walking than when sitting, but when listening to an earpiece, they performed equally well whether walking or sitting. Background visual motion alone, therefore, may make it difficult to use an HMD or may simply exacerbate problems caused by sound motion.

Experiment and Hypotheses

We developed a novel approach to explore the relative contributions of sound motion and background visual motion with Thompson and Sanderson's (2008) results. In the current experiment, a partial replication of Thompson and Sanderson, participants completed the mismatch task while they were walking or sitting and hearing mismatch sounds in free field. We investigated the free-field condition because Thompson and Sanderson's results indicated that it was only when participants used free-field sound that their performance worsened when they were walking. A further issue is that walking involves workload associated with maintaining posture and maintaining visual orientation (e.g., Lajoie, Teasdale, Bard, & Fleury, 1993; Sparrow, Bradshaw, Lamoureux, & Tirosh, 2002).

Accordingly, to remove any workload associated with walking itself and to achieve an independent manipulation of sound motion and background visual motion, participants performed the mismatch task while seated inside a mobile room (see Figure 2). The walls of the mobile room could be rotated around a participant from a ceiling-mounted pivot point to replicate the horizontal flow of background visual motion experienced by walking participants when they turn. Using the mobile room frame, we could rotate the free-field sound source around the participant to create relative sound motion similar to that experienced by the participant when walking. Finally, both the walls and sound source could be rotated simultaneously using the mobile room.

The first hypothesis (Hypothesis 1) is that participants' mismatch task accuracy in the real room will be greater when they are sitting than when walking, as found by Thompson and Sanderson (2008) for performance with free-field sound. Replicating this finding would provide a basis for testing predictions with the mobile room.

The second hypothesis (Hypothesis 2) relates to whether sound motion is sufficient to produce worse mismatch task accuracy. Compared with the condition in which participants are sitting in the real room, participants' accuracy at the mismatch task in the mobile room may be worse with sound motion alone (Hypothesis 2a). Alternatively, participants' accuracy may be worse only when there is both sound motion and background visual motion (Hypothesis 2b).

The third hypothesis (Hypothesis 3) is that compared with when they are sitting in the real room, participants will show worse mismatch task accuracy when they are sitting in the mobile room and there is background visual motion alone. Background visual motion will enhance eye reflexes, making it difficult for participants to monitor the HMD.

Figure 2. Mobile room walls (A) attached for SitBothMove condition and (B) detached for SitSoundMoves condition. In both examples, the speaker is fixed to the rotating frame and so is mobile.

METHOD

Participants

The study was approved by the University of Queensland's School of Psychology Ethical Review Committee. Participants were students who participated either to earn credit for a 1st-year psychology course or for monetary compensation (AUD $20). Participants were 8 males and 18 females, with ages ranging from 18 to 35 years (M = 22.5, SD = 4.1 years). All gave written informed consent and reported normal or corrected-to-normal vision and normal hearing.

Design

Conditions. The experiment used a repeated-measures design. The independent variable, *motion*, represented the conditions in which the participant did the mismatch task: (a) sitting at a computer desk in the real room (Sit),

(b) walking in the real room (Walk), (c) sitting in the mobile room while both the speaker and walls rotated (Sit-BothMove), (d) sitting in the mobile room while the walls were fixed but the speaker rotated (Sit-SoundMoves), and (e) sitting in the mobile room while the walls rotated but the sound was fixed (Sit-WallsMove) (see Figure 3). While doing the mismatch task, in all conditions, participants also performed a secondary button-press task in which they read a letter on a screen and pressed a button on a button box labeled with the same letter (see section on button-press task for more details).

Counterbalancing. Each participant was tested in all five conditions shown in Figure 3. The order of experimental conditions was counterbalanced to remove practice or sequence effects. Five mismatch task scenarios (series of bounce sequences) were presented in the same order across participants. Because the order of experimental conditions was counterbalanced across participants, each experimental condition was observed in combination with all mismatch task scenarios and in all serial positions.

Tasks

Mismatch task. As shown in Figure 1, two soft objects and one hard object moved around a screen, bouncing off each other and off the walls. The surrounding wall was defined as hard. The correct sound for a soft object colliding with any other object was a soft sound, whereas the correct sound for the hard shape hitting the wall was a hard sound. The mismatch task involved detecting violations of the correct behavior, or mismatches, when the visual objects collided and the resulting sound mismatched. For example, a mismatch occurred when there should have been a soft sound but there was actually a hard sound or vice versa. Participants kept a silent mental count of the number of times the visual and auditory behavior of the objects mismatched.

Mismatch task scenarios. Thompson and Sanderson (2008) found that the effect of walking with free-field sound was somewhat stronger when approximately 11% of bounces were mismatches. Accordingly, in this experiment, the number of mismatches in each scenario was 34, 36, 37, 37, and 38, representing 10.8%, 10.6%, 11.3%, 11.3%, and 11.9% of total bounces, respectively. As in the earlier study, the inter-mismatch interval was always greater than 3 s, and bounces occurred on average every 0.73 s. Each scenario lasted 4 min, and the total number of bounces per scenario ranged from 315 to 341.

Mismatch task hardware. The HMD was a Microvision Nomad™ ND2000 with a single optical see-through monocle (800 × 600 pixels). It was connected to a Sony™ U50 tablet computer, which ran the mismatch task software. Mismatch task objects were 80 × 80 pixels and moved at 150 pixels/s. Bounce sounds generated from the U50 were sent to a loudspeaker (Sony SRS-A27 Active Speaker System) via a wireless transmitter (Sony URX-P1/UTX-B1). The sound pressure level (SPL) of the sounds, measured from the center of the experiment room, was 78 dBA max.

Button-press task. As in Thompson and Sanderson's (2008) study, the button-press task ensured that participants would move around the room in the walking condition. Participants also completed the button-press task in all other conditions for control purposes. Four button boxes labeled *A*, *B*, *C*, and *D* were placed at table height in each corner of the room for the Walk condition and in a similar configuration in front of the participant for the various Sit conditions (see Figures 2 and 3). In front of the participant on a computer screen, a large letter indicated the button box to press. The letter was selected quasirandomly from the set (*A*, *B*, *C*, *D*) every 8 s and was displayed until the next letter was displayed. The participant pressed the corresponding button box. A notification sound alerted participants when a new letter was displayed. The notification sound came from a different location and had a different acoustic profile than did the mismatch task sounds.

Button-press task hardware and software. The button-press task was controlled with E-Prime software (Psychology Software Tools, Pittsburgh, PA). In the Sit and Walk conditions, the letters were displayed on a 17-in. LCD display. The height of the letters subtended approximately 24° of visual angle when participants were seated and approximately 7° when participants were standing in the middle of the room. In the

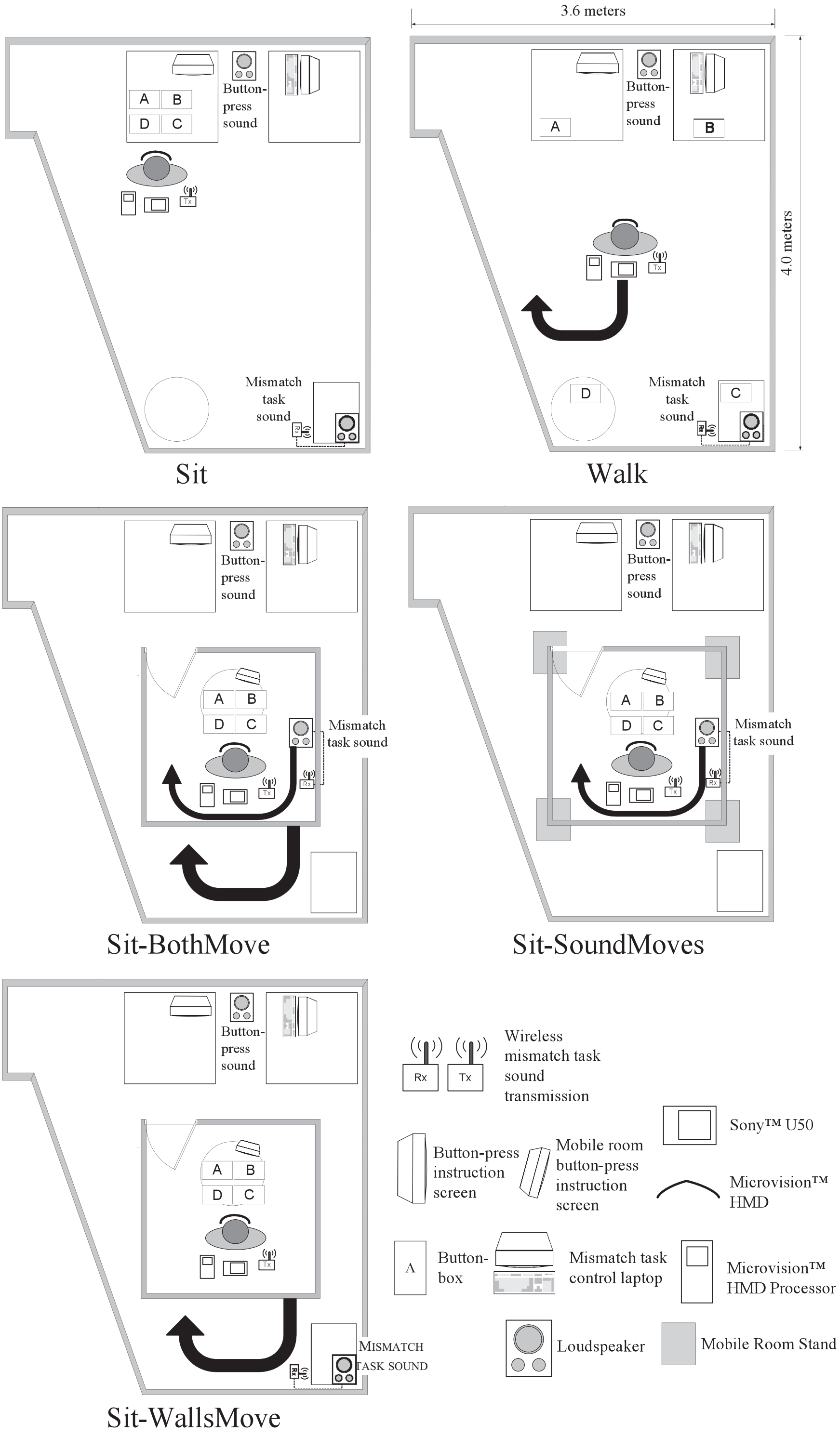

Figure 3. Schematic of different experimental conditions. Arrows indicate one of the two possible directions of motion.

mobile room conditions, the letter was displayed on a 4-in. LCD display (Sony Video Walkman GV-D800E). The height of the letters on the 4-in. display subtended approximately 10° of visual angle when participants were seated inside the mobile room. The push buttons each had a 4.5-cm flat top and so were easy to locate and press with minimal visual guidance. The SPL of the notification sound from the button-press task speaker was approximately 64 dBA max.

Subjective Self-Motion Questionnaire. To assess any subjective self-motion effects caused by the rotating mobile room, participants were asked, “At any stage, did you experience a sense of being in motion?” Participants responded by marking their answers on a scale from 1 (“I felt totally motionless”) to 7 (“I felt like I was in complete motion”).

Apparatus

Layout. Detailed experimental layouts for the five experimental conditions are shown in Figure 3. In the Walk condition, each button box was placed in a different corner of the real room so that participants had to walk around to press the buttons. In the Sit condition, participants sat at a computer desk in the real room with the four button boxes in front of them. In the other three conditions, participants sat inside the mobile room at a desk, on which the button boxes were placed in the same arrangement as for the Sit condition (see Figures 2 and 3).

In the Sit, Walk, and Sit-WallsMove conditions, sounds from the mismatch task came from a speaker in the corner of the test room. In the Sit-BothMove and Sit-SoundMoves conditions, sounds from the mismatch task came from a speaker mounted to the mobile room ceiling frame.

The mobile room. The mobile room was constructed to replicate the relative motion of free-field sound and the dynamic visual background that a participant would experience while turning but to do so when the participant was actually seated. The mobile room was suspended from the laboratory ceiling, where it could be freely rotated about its central, vertical axis around a seated participant (see Figure 2a). Each wall of the mobile room was 1,600 mm × 1,200 mm (length × height). The mobile room wall rested approximately 900 mm from a seated participant’s eyes, compared with a distance of approximately 3 m from the real room front wall to the eyes of a participant standing equidistant from the button boxes.

In the conditions in which the mobile room walls rotated, Sit-BothMove and Sit-WallsMove, the walls were attached to the rotating frame (see Figure 2a). In the condition in which the mobile room walls remained fixed, Sit-SoundMoves, the walls were detached from the rotating frame and placed on 300-mm-high supports (see Figure 2b). For visual consistency between conditions, photos on the walls of the mobile room reproduced the view that each wall of the real room subtended on the retina of a viewer sitting in the middle of the room (see Figure 2a).

Digital photos were taken of the required area of each wall of the real room. The digital photos were printed at a size that preserved the visual angle of the view of the real walls when projected onto the mobile room walls and were then fixed to the mobile room walls (see Figure 2a).

The rotating frame. For conditions Sit-BothMove and Sit-SoundMoves, the speaker and wireless microphone receiver (see button-press task hardware and software section) were attached to the rotating mechanism. When the mobile room walls were attached to the rotating frame, the midpoint of the height of the walls was approximately 1,200 mm, equal to the average sitting eye height from the ground for men and women combined (Pheasant & Haslegrave, 2006).

Rotating the mobile room walls and speaker. The ceiling frame of the mobile room (and consequently the wall and/or speaker) was rotated by the experimenter to approximate the horizontal motion that a participant would experience if he or she were turning toward and away from a button box. For example, if a walking participant initially turned 60° to the right to walk toward button box *B*, the background would initially appear to move to the left by 60°, then to the right by 60°, when the participant turned in the opposite direction to face the front of the room. The directions of rotations of the mobile room were consistent patterns according to the presented letter in the button-press task and were performed in a standardized fashion. All rotations made toward and away from the corresponding button box location were completed before the next letter was presented (within 8 s).

Procedure

Participants adjusted the focus on the HMD using the hyperopic procedure so that the HMD display was approximately the same focal distance as a wall viewed from 1.5 m (Behar, Wiley, Levine, Rash, & Walsh, 1990). Participants learned components of the mismatch task and then the full mismatch task, and they completed two full practice trials as per Thompson and Sanderson's (2008) study, each lasting 4 min. Walk was selected as a practice condition because Thompson and Sanderson found Walk was the more difficult condition. A mobile room condition was given as practice so participants could experience sitting inside the mobile room. Specifically, Sit-SoundMoves was selected for training and was a conservative choice because the literature suggested that the moving sound source alone might interfere with the mismatch task more than would the moving background (e.g., Zampini et al., 2005; Laramee & Ware, 2002).

In the Walk practice trial, the experimenter coached participants to face each button when pressing it. During the practice trials and throughout the experiment, participants were told that the mismatch task was the most important task but to do the button-press task as efficiently as possible.

Dependent Variables

The main dependent variable in each condition was the number of mismatches that participants reported divided by the correct number. Other dependent variables included the accuracy and latency of participants' performance with the secondary button-press task and participants' responses to the postcondition questionnaire about subjective self-motion.

RESULTS

Mismatch Task Accuracy

Mismatch accuracies were calculated by dividing the number of mismatches that participants reported by the actual number of mismatches per scenario and expressing the result as a percentage. A one-way repeated-measures ANOVA was conducted on mismatch count data with the within-subjects factor of motion (five levels: Sit, Walk, Sit-BothMove, Sit-SoundMoves, and Sit-WallsMove).

Results are shown in Figure 4 and Table 1. The significant main effect of motion showed that participants' mismatch accuracy varied across conditions, $F(4, 100) = 3.13$, $MSE = 0.01$, $p = .017$. Two-tailed Fisher's LSD post hoc tests were used to determine differences between means (Saville, 1990).

Hypothesis 1 was that participants' mismatch accuracy would be better in the Sit condition than in the Walk condition. As predicted, participants counted mismatches more accurately in the Sit condition ($M = 79\%$, $SD = 21\%$) than in the Walk condition ($M = 73\%$, $SD = 22\%$), $p = .032$. (In Thompson and Sanderson, 2008, accuracy for the Sit condition was 87% and for the Walk condition was 75%.)

Hypothesis 2 tested the role of sound motion and whether, compared with the Sit condition, participants' mismatch accuracy would be worse in the Sit-SoundMoves condition or only in the Sit-BothMove condition. In partial support of Hypothesis 2b, participants showed a trend to identify mismatches less accurately in the Sit-BothMove condition ($M = 74\%$, $SD = 20\%$), $p = .058$, but no less accurately in the Sit-SoundMoves condition ($M = 81\%$, $SD = 20\%$), $p = .461$. Therefore, even when the location of the sound source changed somewhat unpredictably, participants' accuracy at the mismatch task was not impaired. Hypothesis 3 tested the role of background visual motion and whether, compared with the Sit condition, participants' mismatch accuracy would be worse in the Sit-WallsMove condition. This hypothesis was not supported. Participants' mismatch accuracy in the Sit-WallsMove condition was not significantly worse than in the Sit condition ($M = 76\%$, $SD = 24\%$), $p = .223$.

There were other significant results that were not directly predicted (see Table 1). Participants counted mismatches significantly more accurately in the Sit-SoundMoves condition than in the Walk condition, $p = .004$. The Walk and Sit-SoundMoves conditions were also used as practice conditions, so both might be expected to have benefited from the practice, but this pattern of results removes any

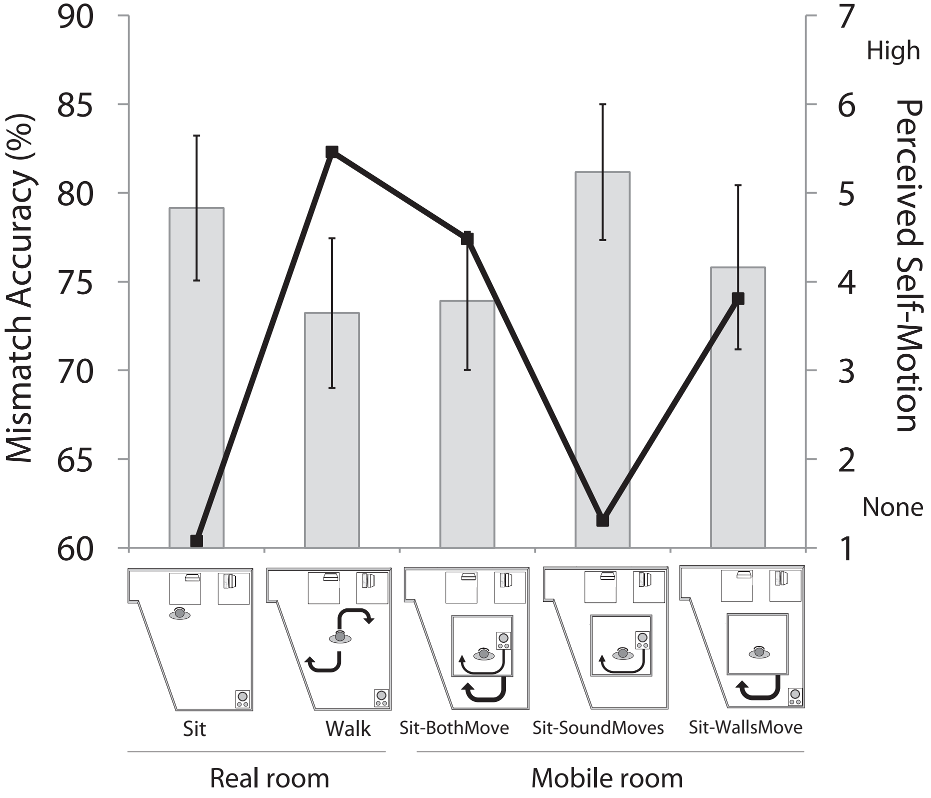

Figure 4. Mismatch accuracy (gray bars, left vertical axis) and levels of perceived self-motion (black line, right vertical axis) across conditions of motion. Mismatch accuracy error bars denote one standard error above and below the mean. Error bars for perceived self-motion have been omitted for clarity but were all less than .5 of a unit.

TABLE 1: Means, Standard Deviations, and LSD *p* Values for Comparisons of All Mismatch Accuracies

	Sit	Walk	Sit-BothMove	Sit-SoundMoves	Sit-WallsMove
M (*SD*)	79% (21%)	73% (22%)	74% (20%)	81% (20%)	76% (24%)
Sit	—				
Walk	*p* = .032**	—			
Sit-BothMove	*p* = .058*	*p* = .802	—		
Sit-SoundMoves	*p* = .461	*p* = .004***	*p* = .009***	—	
Sit-WallsMove	*p* = .223	*p* = .346	*p* = .489	*p* = .052*	—

**p* < .10. ***p* < .05. ****p* < .01.

concern about practice effects. Participants also counted mismatches significantly more accurately in the Sit-SoundMoves condition than in the Sit-BothMove condition, $p = .009$, and they showed a trend to count mismatches more accurately than in the Sit-WallsMove condition, $p = .052$. No other comparisons were significant.

Button-Press Task

Results for the button-press task established that there were no trade-offs between the mismatch task and the button-press task across conditions and no speed–accuracy trade-offs within conditions.

Time taken to push button boxes. A repeated-measures ANOVA was conducted on participants' latency at pushing button boxes, with the five within-subjects levels of motion and four levels of button (*A*, *B*, *C*, and *D*).

There was a significant main effect of motion $F(4, 100) = 195.26$, $MSE < 0.001$, $p < .001$, a significant main effect of button, $F(3, 75) = 23.34$, $MSE < 0.001$, $p < .001$, and significant interaction between motion and button, $F(12, 300) = 2.37$, $MSE < 0.001$, $p = .006$. A Tukey HSD revealed that as expected, participants took longer to push every button when they were walking (M = 4,279 ms, SD = 814 ms) than in the other four conditions, all ps < .001. There were no differences between the nonwalking conditions (range = 1,853 ms to 1,994 ms), suggesting that there was no tradeoff between the mismatch task and the button-press task. Push-button responses were slightly, but very reliably, faster to *A* and *B* than to *C* and *D*, especially for participants who were walking.

Button-press task accuracy. Participants' accuracy at pressing button boxes was better than 90% in all conditions. Accuracy was not normally distributed, and many conditions were performed at 100% accuracy by many participants. A nonparametric Friedman ANOVA indicated that there was no difference in how accurately participants pressed buttons across conditions, $\chi^2(19, N = 26) = 18.35$, $p = .50$, *ns*.

Subjective Self-Motion

Participants rated perceived self-motion on a scale from 1 to 7, with higher ratings representing greater levels of perceived self-motion. Means for responses across conditions are shown in Figure 4. A Friedman ANOVA revealed that means differed across conditions, $\chi^2(4, N = 26) = 67.98$, $p < .001$. Wilcoxon matched pairs tests indicated that participants' experience of being in motion was greater in the Walk condition than in either Sit ($p < .001$), Sit-SoundMoves ($p < .001$), or Sit-WallsMove ($p = .018$) conditions. There was no difference in participants' experience of being in motion between the Walk and Sit-BothMove conditions, $p = .225$, *ns*. Compared with the Sit and Sit-SoundMoves conditions, participants' experience of being in motion was greater in the Sit-BothMove and Sit-WallsMove conditions, all ps < .001. There was a negative correlation between the overall means for self-motion and for mismatch accuracy, $r(3) = -.96$, $p = .009$.

DISCUSSION

Our aim was to establish the individual and combined effects of a moving sound source and a moving visual background on people's ability to integrate vision on an HMD with sound delivered via a free-field speaker. As predicted in the first hypothesis, participants' mismatch accuracy in the real room was better when they were sitting than when they were walking (see Figure 4). This finding corroborates the results of Thompson and Sanderson (2008), supporting the generality of their finding, although accuracy for the Sit condition was not quite as high as in their study. The present finding also provides a firm basis from which to investigate the effect of sound motion and background visual motion on mismatch accuracy in the mobile room when the workload of walking is removed.

As predicted in Hypothesis 2b, when participants are in the mobile room, sound motion alone is insufficient to reduce their mismatch accuracy significantly, compared with when they were sitting in the real room. Instead, both background visual motion and sound motion were required to reduce participants' mismatch accuracy to a degree that approached significance. In addition, mismatch accuracy was very similar when participants were walking in the real room versus when they were experiencing both background visual motion and sound motion in the mobile room. Human movement studies suggest that walking imposes workload (Lajoie et al., 1993; Sparrow et al., 2002), but the present results suggest that the combination of sound motion and background visual motion, rather than whether they are walking, accounts for differences in participants' mismatch accuracy.

Contrary to Hypothesis 3, when participants were sitting in the mobile room and experiencing only background visual motion, their mismatch accuracy was not significantly worse than when they were sitting in the real room. Because our manipulation of background visual motion was similar to other eye reflex experiments (Williams et al., 2006), we had expected that background visual motion would make it more difficult for participants to track HMD-based objects. The lack of difference is more like Thompson and Sanderson's (2008) findings for sound delivered by earpiece than for sound delivered in free field. Nonetheless, in our results, mismatch accuracy was marginally worse when there was background visual motion than when there was sound motion alone. Moreover, the combination of background visual motion and sound motion seemed to play a crucial role in how well participants could detect mismatches, as will be discussed in more detail in the next sections.

Mismatch Accuracy and Multisensory Integration Research

Although basic multisensory integration research provided a basis for predictions, the results of the present experiment cannot be explained by multisensory integration research alone. Previous multisensory integration research has found that as visual and auditory stimuli become spatially separated, the likelihood decreases that they will be integrated (Jack & Thurlow, 1973; Slutsky & Recanzone, 2001; Thurlow & Jack, 1973). Similarly, it has been shown that when the spatial relationships of simultaneously presented auditory and visual stimuli are inconsistent and unpredictable, people sometimes incorrectly report that the stimuli have occurred in sequence (Slutsky & Recanzone, 2001; Zampini et al., 2005).

In the current experiment, however, when participants were sitting in the mobile room and experiencing sound motion, their mismatch accuracy was no different than when they were sitting in the real room and there was no sound motion. This result means that even when the location of the sound source changed somewhat unpredictably, participants' accuracy at the mismatch task was not impaired. Thompson and Sanderson (2008) suggested that the changing spatial relationships between the source of the bounce sounds versus the visual stimuli on the HMD could affect participants' perception of the order in which stimuli occurred, but the current results rule out this explanation.

Groh and Werner-Reiss (2002) have shown the importance of reconciling visual and auditory "reference frames." In the unusual circumstances of our experimental conditions, the process by which participants reconcile the auditory and visual reference frames could have been compromised, but this appears not to have happened. For example, the Walk and Sit-BothMove conditions lead to similar performance, whereas the way the auditory versus visual reference frames have to be reconciled differs in each condition. Multiple resource theory also might appear to offer insights, but it does not distinguish resource implications of the location or motion of sounds (Wickens, 2002). Moreover, although locomotion has possible resource implications, it does not affect our results.

Combining Background Visual Motion and Sound Motion

Our results suggest that background visual motion plays a crucial role in how well people can integrate sound with vision on an HMD but that background visual motion and sound motion together lead to the strongest decrease in participants' performance. This is consistent with Thompson and Sanderson's (2008) findings. As shown in Figure 4, participants experienced subjective self-motion most strongly with both background visual motion and sound motion. Independent research corroborates the fact that moving sounds can enhance visually induced subjective self-motion, or vection, although the effect is weaker than for visual motion (Riecke, Väljamäe, & Schulte-Pelkum, 2009). Eye movements are typically made in response to self-motion (e.g., Leigh & Zee, 1999).

Given the above factors, participants may have found it harder to monitor the HMD because background visual motion created visual interference (Laramee & Ware, 2002). As a result, participants may have had to suppress eye reflexes, such as OKN, quite forcefully, as has been found in vision research more generally (Leigh & Zee, 1999; Paige, Telford, Seidman, & Barnes, 1998; Schweigart, Mergner & Barnes, 1999). In addition, the role of sound motion (Riecke et al., 2009) in enhancing vection, plus the additional cognitive load of covertly attending to moving sounds, may have made suppression of eye reflexes even more difficult (Colonius & Arndt, 2001; Hutton & Tegally, 2005; Williams et al., 2006) so that participants occasionally lost fixation on the HMD when a mismatch occurred. However, we cannot say whether the possible role of eye reflexes explains all failures to count mismatches or just some.

Limitations and Future Directions

Possible limitations of our study are as follows. First, eye reflexes made in response to background visual motion and self-motion were not directly tested, and we do not know how much of the performance decrement they accounted for. Future studies should assess directly the role of eye reflexes.

Second, it has been suggested by reviewers of Thompson and Sanderson (2008) and related works that requiring participants to maintain a mental count of mismatches may induce workload that causes the present results. However, positing interactions between the workload of maintaining a mental count and the experimental conditions does not offer a parsimonious explanation for either the Thompson and Sanderson (2008) results or the present results.

Third, the mental-count dependent variable is relatively insensitive. Another study has introduced a clicker that allows participants to register mismatches when they detect them (Thompson, Tear, & Sanderson, 2010 [this issue]). Although informative signal detection measures can be calculated, the clicker appears to increase the motor coordination demands of the task sufficiently to bring other factors to bear on performance that change the pattern of results. Further investigation is required to establish the sensitivity of our findings to the response method.

There are two major directions for future research. First, researchers could examine the necessary and sufficient conditions for a drop in performance on tasks involving multiple modalities when self-motion is involved. As noted in the Introduction, Väljamäe (2009) highlights the need for future research to understand more fully how optokinetic stimulation, vection, and auditory perception may interact. As also noted, Teramoto et al. (2004, 2008) have shown that under conditions of self-motion, people misjudge the temporal order of visual events, of sounds, and of tactile stimuli, but they have not examined the effect of self-motion on multisensory integration across any of these modalities. Further research is needed with procedures such as theirs to corroborate our findings and to examine the role of eye reflexes in such phenomena. In an experimental design similar to Williams et al. (2006), eye-tracking technology could be used to test whether a moving visual environment affects an HMD user's ability to stay fixated on the HMD and to clarify the role of self-motion.

Second, researchers could explore arrangements of stimuli that represent other situations in which participants are making multisensory judgments. Other tasks involving decision making based on multiple modalities—especially decisions appearing to involve multisensory integration—should be explored, such as "way finding" (Eriksson et al., 2008). Although we have demonstrated that walking drove the effect found by Thompson and Sanderson (2008), it was specifically the background motion and sound motion experienced during walking that affected performance, not walking itself. Walking is just one form of self-motion, and the generality of our findings for further forms of self-motion should be explored, such as the self-motion experienced when controlling or riding in a vehicle.

CONCLUSION

Although advanced displays often confer performance advantages (Sanderson et al., 2008), the present results show that the attentional demands imposed by such technologies may sometimes have detrimental effects. Covertly monitoring a moving auditory source may detract from people's ability to suppress eye reflexes and stay fixated on an HMD. Commercially available speakers are a cost-effective way of delivering sound information to the HMD user. However, if HMD users are required to integrate auditory and visual information, their ability to do so may suffer when there is both strong background visual motion and sound motion, as occurs when the HMD user is walking around or controlling a vehicle. We therefore encourage researchers to include self-motion as a manipulation that might introduce greater representativeness to studies that could generalize to dynamic work environments.

ACKNOWLEDGMENTS

The present research was supported by Australian Research Council Discovery Grant DP0559504. We thank Microvision, and especially Andrew Rosen, for kindly arranging a loan of a Microvision Nomad ND 2000 HMD that allowed us to complete the research. Finally, we acknowledge the help of members of the Cognitive Engineering Research Group and Jason Tangen at The University of Queensland for support throughout the research.

REFERENCES

Behar, I., Wiley, R. W., Levine, R. R., Rash, C. E., & Walsh, D. J. (1990). *Visual survey of Apache aviators (VISAA)* (No. ADA230201). Fort Rucker, AL: U.S. Army Aeromedical Research Laboratory.

Bertelson, P., & de Gelder, B. (2004). The psychology of multimodal perception. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 141–177). New York, NY: Oxford University Press.

Colonius, H., & Arndt, P. (2001). A two-stage model for visual-auditory interaction in saccadic latencies. *Perception & Psychophysics*, *63*, 126–147.

Eriksson, L., Berglund, A., Willen, B., Svensson, J., Petterstedt, M., Carlander, O., . . . Allerbo, G. (2008). On visual, vibrotactile, and 3D audio directional cues for dismounted soldier waypoint naviation. In *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting* (pp. 1282–1286). Santa Monica, CA: Human Factors and Ergonomics Society.

Groh, J. M., & Werner-Reiss, U. (2002). Visual and auditory integration. In V. S. Ramachandran (Ed.), *Encyclopaedia of the human brain* (Vol. 4, pp. 739–752). San Diego, CA: Academic Press.

Hutton, S. B., & Tegally, D. (2005). The effects of dividing attention on smooth pursuit eye tracking. *Experimental Brain Research*, *163*, 306–313.

Jack, C. E., & Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. *Perceptual and Motor Skills*, *37*, 967–979.

Lajoie, Y., Teasdale, N., Bard, C., & Fleury, M. (1993). Attentional demands for static and dynamic equilibrium. *Experimental Brain Research*, *97*, 139–144.

Laramee, R. S., & Ware, C. (2002). Rivalry and interference with a head-mounted display. *ACM Transactions on Computer-Human Interaction*, *9*, 238–251.

Leigh, R., & Zee, D. (1999). *The neurology of eye movements* (3rd ed.). New York, NY: Oxford University Press.

Liu, D., Jenkins, S., & Sanderson, P. M. (2009). Patient monitoring with head-mounted displays. *Current Opinion in Anaesthesiology*, *22*, 796–803.

Liu, D., Jenkins, S., Sanderson, P., Watson, M., Leane, T., Kruys, A., & Russell, W. J. (2009). Monitoring with head-mounted displays: Performance and safety in a full-scale simulator and part-task trainer. *Anesthesia & Analgesia*, *109*, 1135–1146.

Paige, G. D., Telford, L., Seidman, S. H., & Barnes, G. R. (1998). Human vestibuloocular reflex and its interactions with vision and fixation distance during linear and angular head movement. *Journal of Neurophysiology*, *80*, 2391–2404.

Pheasant, S., & Haslegrave, C. M. (2006). *Bodyspace: Anthropometry, ergonomics and the design of work* (3rd ed.). London, England: Taylor and Francis.

Riecke, B. E., Väljamäe, A., & Schulte-Pelkum, J. (2009). Moving sounds enhance the visually-induced self-motion illustion (circular vection) in virtual reality. *ACM Transactions on Applied Perception*, *6*(2), 7:1–7:27.

Sanderson, P. M., Watson, M. O., Russell, W. J., Jenkins, S., Liu, D., Green, N., . . . Krupenia, S. (2008). Advanced auditory displays and head-mounted displays: Advantages and disadvantages for monitoring by the distracted anesthesiologist. *Anesthesia & Analgesia*, *106*, 1787–1797.

Saville, D. J. (1990). Multiple comparison procedures: The practical solution. *American Statistician*, *44*, 174–180.

Schweigart, G., Mergner, T., & Barnes, G. (1999). Eye movements during combined pursuit, optokinetic and vestibular stimulation in macaque monkey. *Experimental Brain Research*, *127*, 54–66.

Sharpe, J. A., & Johnston, J. L. (1993). The vestibulo-ocular reflex: Clinical, anatomic, and physiologic correlates. In J. A. Sharpe & H. O. Barber (Eds.), *The vestibulo-ocular reflex and vertigo* (pp. 15–39). New York, NY: Raven Press.

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*, *12*, 7–10.

Sparrow, W. A., Bradshaw, E. J., Lamoureux, E., & Tirosh, O. (2002). Ageing effects on the attention demands of walking. *Human Movement Science*, *21*, 961–972.

Teramoto, W., Watanabe, H., & Umemura, H. (2008). Change of temporal-order judgment of sounds during long-lasting exposure to large-field visual motion. *Perception*, *37*, 1649–1666.

Teramoto, W., Watanabe, H., & Umemura, H., Matsuoka, K., & Kita, S. (2004). Judgment biases of temporal order during apparent self-motion. *IEICE Transactions on Information and Systems*, *E87D*, 1466–1476.

Thompson, M. B., & Sanderson, P. M. (2008). Multisensory integration with a head-mounted display: Sound delivery and self-motion. *Human Factors*, *50*, 789–800.

Thompson, M. B., Tear, M. J., & Sanderson, P. M. (2010). Multisensory integration with a head-mounted display: Role of response modality. *Human Factors*, *52*, 92-104.

Thurlow, W. R., & Jack, C. E. (1973). A study of certain determinants of the ventriloquism effect. *Perceptual and Motor Skills*, *36*, 1171–1184.

Väljamäe, A. (2009). Auditorily-induced illusory self-motion: A review. *Brain Research Reviews*, *61*, 240–255.

Veltman, J. A., Oving, A. B., & Bronkhorst, A. W. (2004). 3-D audio in the fighter cockpit improves task performance. *International Journal of Aviation Psychology*, *14*, 239–256.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*, 159–177.

Williams, I., Mulhall, L., Mattingley, J., Lueck, C., & Abel, L. (2006). Optokinetic nystagmus as an assessment of visual attention to divided stimuli. *Journal of Clinical Neuroscience*, *13*, 828–833.

Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audiovisual simultaneity judgments. *Perception and Psychophysics*, *67*, 531–544.

Zotkin, D. N., Duraiswami, R., Grassi, E., & Gumerov, N. A. (2006). Fast head-related transfer function measurement via reciprocity. *Acoustical Society of America*, *120*, 2202–2215.

William J. Harrison obtained his BA (Hons) in psychology in 2008 from The University of Queensland in St. Lucia, Queensland, Australia, where he is currently a PhD student.

Matthew B. Thompson received a dual BSc (Hons) in psychology and bachelor of information technology in 2007 from The University of Queensland in St. Lucia, Queensland, Australia, where he is currently a PhD student.

Penelope M. Sanderson obtained her PhD in engineering psychology from the University of Toronto in 1985. She is a professor of cognitive engineering and human factors at The University of Queensland in St. Lucia, Queensland, Australia, with appointments in the Schools of Psychology, of Medicine, and of Information Technology and Electrical Engineering.

Date received: June 11, 2009
Date accepted: February 22, 2010