Multisensory Integration With a Head-Mounted Display: Sound Delivery and Self-Motion

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

Objective: We tested whether the method of sound delivery affects people's ability to integrate information from multiple modalities when they are walking and using a head-mounted display (HMD). **Background:** HMDs increasingly support mobile work. Human operators may benefit from auditory support when using an HMD. However, it is unclear whether sound is better delivered publicly in free field or privately via earpiece and what the effect of walking is. **Method:** Participants identified mismatches between sounds and visual information on an HMD. Participants heard the sounds via either earpiece or free field while they either sat or walked about the test room. **Results:** When using an earpiece, participants performed the mismatch task equally well whether sitting or walking, but when using free-field sound, participants performed the task significantly worse when walking than when sitting ($p = .006$). **Conclusion:** The worse performance for participants using free-field sound while walking may relate to spatial and motion inconsistencies between visual events on the head-referenced HMD and auditory events from world-referenced speakers. Researchers should more frequently examine the effect of self-motion on people's ability to perform various multisensory tasks. **Application:** When multisensory integration tasks are performed with an HMD and free-field delivery of sound, as may happen in medicine, transportation, or industry, performance may suffer when the relative location of sound changes as the user moves.

INTRODUCTION

In the research reported here, we tested whether people's ability to integrate information from head-mounted displays (HMDs) and auditory displays depends on how the auditory information is presented and whether people move about when working. HMDs display information over the forward field of view, thus making information continuously available to a mobile user (Laramee & Ware, 2002; Patterson, Winterbottom, & Pierce, 2006; Perrott, Cisneros, McKinley, & D'Angelo, 1996; Yeh, Merlo, Wickens, & Brandenburg, 2003). Advanced auditory displays such as auditory icons, earcons, or sonification can be useful when vision is unavailable, inadequate, or overloaded to help a user maintain peripheral awareness and to draw the user's attention quickly to important state changes (Barrass & Kramer, 1999; Brewster, 1994; Kramer, 1994; Walker & Kramer, 2004; Watson & Sanderson, 2004, 2007). How do such displays work together as a multimodal environment for the mobile user?

Many jobs require people to move about. Anesthesiologists walk between different locations in the operating theater. Workers in freight warehouses must move to different locations. Technicians maintaining large vehicles or plants often work in continually changing orientations. In such cases HMDs can be useful (Ockerman & Pritchett, 1998; Ong, Yuan, & Nee, 2008; Ormerod, Ross, & Naluai-Cecchini, 2003), but it is unclear how complementary auditory information should be presented. In some contexts it can require considerable technical sophistication and expense to create the impression that visual and auditory information about an object or process come from precisely the same spatial location, as they would in the natural environment. In the operating theater, for example, auditory information is almost always presented in free field, but listening via an earpiece may be more practical as more advanced

Address correspondence to Penelope Sanderson, ARC Key Centre for Human Factors, The University of Queensland, St Lucia, QLD, Australia 4072; psanderson@itee.uq.edu.au. *HUMAN FACTORS,* Vol. 50, No. 5, October 2008, pp. 789–800. DOI 10.1518/001872008X312323. Copyright © 2008, Human Factors and Ergonomics Society.

auditory displays are introduced and if health care workers move about (Sanderson et al., 2008). Mobility therefore poses challenges for how multimodal information is best presented.

Although there is much research on HMDs and on auditory displays alone, there is less research on how they might work together when people are mobile. As the wearer moves, an HMD will be viewed against different backgrounds. An auditory display will be heard in different acoustic or social contexts, in which different sound delivery methods may be effective and appropriate. Although social and communication issues have been studied with the mobile user (Roibas, Geerts, Furtado, & Calvi, 2008), perceptual issues have received relatively little attention.

We present an initial study on the role of both self-motion and sound delivery method on people's ability to perform multisensory integration of auditory information with visual information from an HMD. Given the relative lack of prior research in this area, we indicate the gaps in the applied and basic literature and describe the experimental task we have developed to initiate research on these issues.

Sound Delivery and Self-Motion

The human factors and human movement literature does not clearly indicate how different methods of sound delivery (earpiece, headphone, free field) and self-motion (walking, standing, sitting) might affect people's ability to perform multisensory integration with an HMD. In the few studies comparing sound delivery methods, researchers have studied preferences for headphone versus free-field delivery of news stories (Kallinen & Ravaja, 2007), the effect of headphone versus free-field music on attention (Nelson & Nilsson, 1990), and the effect of 2-D versus 3-D headphones on people's ability to use an HMD to navigate in virtual space (Viaud-Delmon, Warusfel, Seguelas, Rio, & Jouvent, 2006). No studies have compared the effectiveness of different methods of sound delivery when a participant is moving.

In other seemingly relevant studies, researchers investigated spatialization of sound to support visual tasks. Participants identify the location of sounds better when sound is delivered through speakers than delivered through headphones (Bolia, D'Angelo, & McKinley, 1999; Perrott et al., 1996). However, participants were seated during such studies rather than walking about.

Further, no studies have directly examined the effect of self-motion on HMD use. Self-motion can introduce attentional demands: Responses to both auditory and visual secondary task probes slow when participants are walking rather than standing or sitting (Lajoie, Teasdale, Bard, & Fleury, 1993; Sparrow, Bradshaw, Lamoureux, & Tirosh, 2002). However, this finding does not indicate whether walking will interfere with the use of an HMD or with some methods of sound delivery more than with others. Laramee and Ware (2002) showed that participants using a transparent HMD found it harder to perform visual tasks against a dynamic visual background, but the participants were seated. In summary, the applied literature does not indicate how sound delivery method or self-motion might affect multisensory integration with an HMD.

Experimental Task

Many different kinds of multisensory tasks emerge when people use an HMD and an auditory display together. For example, people may need to divide attention across modalities or to integrate information arriving in different modalities. However, we found no continuous multisensory integration task of proven sensitivity that would be suitable for our investigation. Therefore we tested the effect of sound delivery and self-motion with a potentially very sensitive multisensory task that we developed, based loosely on the Michotte (1954/1963) launch task– the *mismatch task.* Next we will describe the launch task and the mismatch task so that our performance predictions can be better understood.

In Michotte's (1954/1963) launch task, if one visual object "hits" another object and makes it move within a certain time window, people perceive a cause-effect relationship. An auditory cue at about the same time strengthens the perception of causality and widens the time window within which causality is perceived (Guski & Troje, 2003). In our mismatch task, participants use the HMD to monitor objects moving and bouncing off each other and the walls (see Figure 1). Some of the objects look hard, and others look soft. When objects collide, they make a hard or soft sound that usually matches the way the objects look, but occasionally there is a mismatch. Participants must integrate the visual and auditory information to detect whether it matches or mismatches. Accordingly, factors that promote multisensory integration

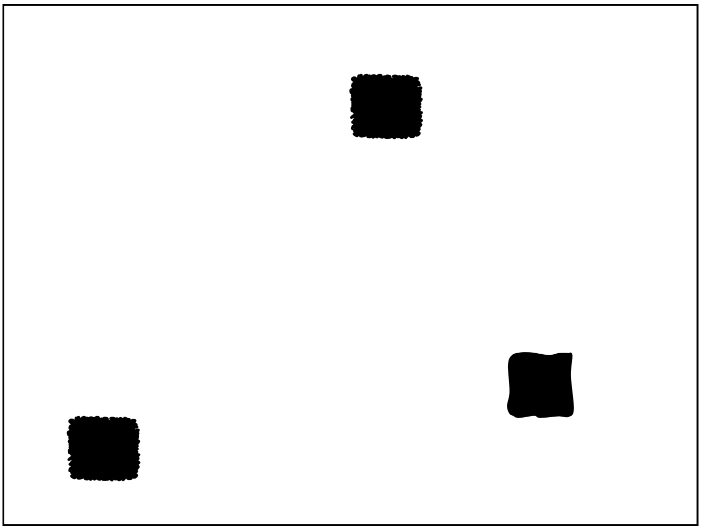

Figure 1. Hard and soft objects as displayed for the mismatch task on the head-mounted display. The hard object is at the bottom right and has smooth, contoured "metal" edges. The soft objects are at the top and bottom left and have textured "woolly" edges. The wall is considered a hard object. Any bounce involving one or more soft objects makes a soft sound, whereas a bounce between two hard objects makes a hard sound.

should improve performance on the mismatch task and factors that challenge multisensory integration should challenge performance.

In our experiment, participants performed the mismatch task while seated or walking and while hearing the sound through a free-field speaker or through an earpiece. They kept count of mismatches in memory. In order to predict whether the method of sound delivery and the participant's motion might promote or challenge multisensory integration, we now turn to the multisensory integration literature.

Multisensory Integration

When an object's properties are not fully specified by any single modality, information from other modalities is used to provide a better representation of the object (Giard & Peronnet, 1999; Newell, 2004; Spence & Driver, 2004; Sumby & Pollack, 1954). For multisensory object integration to occur, information from different modalities should be (a) task relevant, (b) temporally congruent, and (c) spatially congruent (Newell, 2004).

The visual and auditory information in the mismatch task is *task relevant* because information from both modalities is always needed to detect a mismatch. The sound and vision of mismatch task bounces occur sufficiently close in time to seem *temporally congruent,* whether the sound is delivered via earpiece or free field. (Time tolerances can be quite large, especially when cause-effect relationships are suggested between visual events and sounds; see Lewald & Guski, 2004.) However, in the mismatch task the sound and vision are never perfectly *spatially congruent.* Spatial congruence takes several different forms in the mismatch task, and it is challenged more strongly in some conditions than in others.

In the mismatch task, the visual objects on the HMD move around the forward field of view. There is no experimental condition in which the bounce sounds are perfectly spatially congruent with the location of visual bounces. When heard through an earpiece, the sound always comes from the right ear, regardless of whether the participant is seated or walking. In the free-field condition when the participant is seated, the sound comes from a speaker to the right and behind the participant. In all three cases, bounces at different visual locations are matched by sounds coming from a fixed location. The spatial mapping between visual and sound events may not be spatially congruent or compatible, but it is consistent.

Spatial incongruence and incompatibility can affect multisensory integration. Multisensory integration performance suffers when an event's vision and sound come from different spatial locations (e.g., Driver & Spence, 1994). As event sound and vision become more spatially separated, ventriloquism effects lessen (Jack & Thurlow, 1973; Slutsky & Recanzone, 2001; Thurlow & Jack, 1973). Temporal order judgments for lights are slower and less accurate if accompanied by spatially incompatible sound stimuli (Vroomen & Keetels, 2006). In the mismatch task, however, such effects do not differentiate the two earpiece conditions or the free-field condition in which participants are seated.

In contrast, in the free-field condition in which the participant walks around, the spatial mappings not only are incongruent and incompatible, but also are dynamically *inconsistent.* The relative locations of objects on the HMD and their sounds continually change because the relative azimuth of the free-field speaker changes as the participant moves around the test room. For example, a visual bounce in the left visual field may be accompanied by a bounce sound from the left when the participant faces in one direction but from the right when the participant faces in the opposite direction. Motion mappings will also be inconsistent. For example, the visual object may move left but, because of the different directions in which the participant may rotate, the bounce sound may appear to move rightward on one occasion but leftward on another occasion.

Spatial and motion incongruence, incompatibility, and inconsistency may challenge participants' ability to detect bounce mismatches. Vision and sound may not appear to come from the same object because Gestalt principles of proximity and common fate are violated (Radeau, 1994). Participants' ability to judge the sequence in which visual events occur may be challenged because people are worse at judging the order in which two lights illuminate when accompanying sounds move across the visual field than when the sounds are stationary (Vroomen & Keetels, 2006). Moreover, sounds may get captured by the wrong visual events: When simultaneous visual and sound events have different directions of motion, people often report that the sound is moving in the same direction as the visual stimulus (Soto-Faraco, Kingstone, & Spence, 2006). For all these reasons, mismatch performance may be worse when sound is delivered in free field to participants who are walking around the test room.

Hypotheses

The first hypothesis is that participants will count mismatches less accurately when they are walking than when they are sitting, regardless of whether sound is delivered via free field or earpiece. This is due to (a) visual interference from the background (Laramee & Ware, 2002) and (b) general attentional resource competition between tasks (Sparrow et al., 2002).

The second hypothesis is that when participants are sitting they will count mismatches equally well with earpiece and free-field sound, but when participants are walking they will count mismatches less accurately with free-field sound than with an earpiece. For walking participants using free-field sound, there will be incompatibility and inconsistency between sound and vision for the position of stimuli and the direction of their motion (Radeau, 1994; Soto-Faraco et al., 2006; Vroomen & Keetels, 2006).

METHOD

Participants

The study was approved by The University of Queensland's Behavioural and Social Sciences Ethical Review Committee. Participants were 20 first-year psychology students between 17 and 39 years of age ($M = 22.3$, $SD = 5.6$) who earned course credit for participation. All gave written informed consent and reported normal or corrected-to-normal vision and normal hearing.

Design

All participants experienced eight conditions created by crossing movement (walking vs. sitting), sound delivery (free field vs. earpiece), and mismatch rate (low vs. high).

Movement. For the walking condition, a button box was placed at each corner of the test room so that participants had to move around the room to press the buttons (see Figures 2 and 3, top panels). For the sitting condition, participants sat at a table with the four button boxes in front of them, with their head stabilized on a chin rest (see Figures 2 and 3, bottom panels). The chin rest limited motion of the HMD across the background and changes in sound azimuth.

Sound delivery. In the free-field condition, object bounce sounds came from a speaker in the corner of the test room. In the earpiece condition, the sounds came from an earpiece in the participant's right ear.

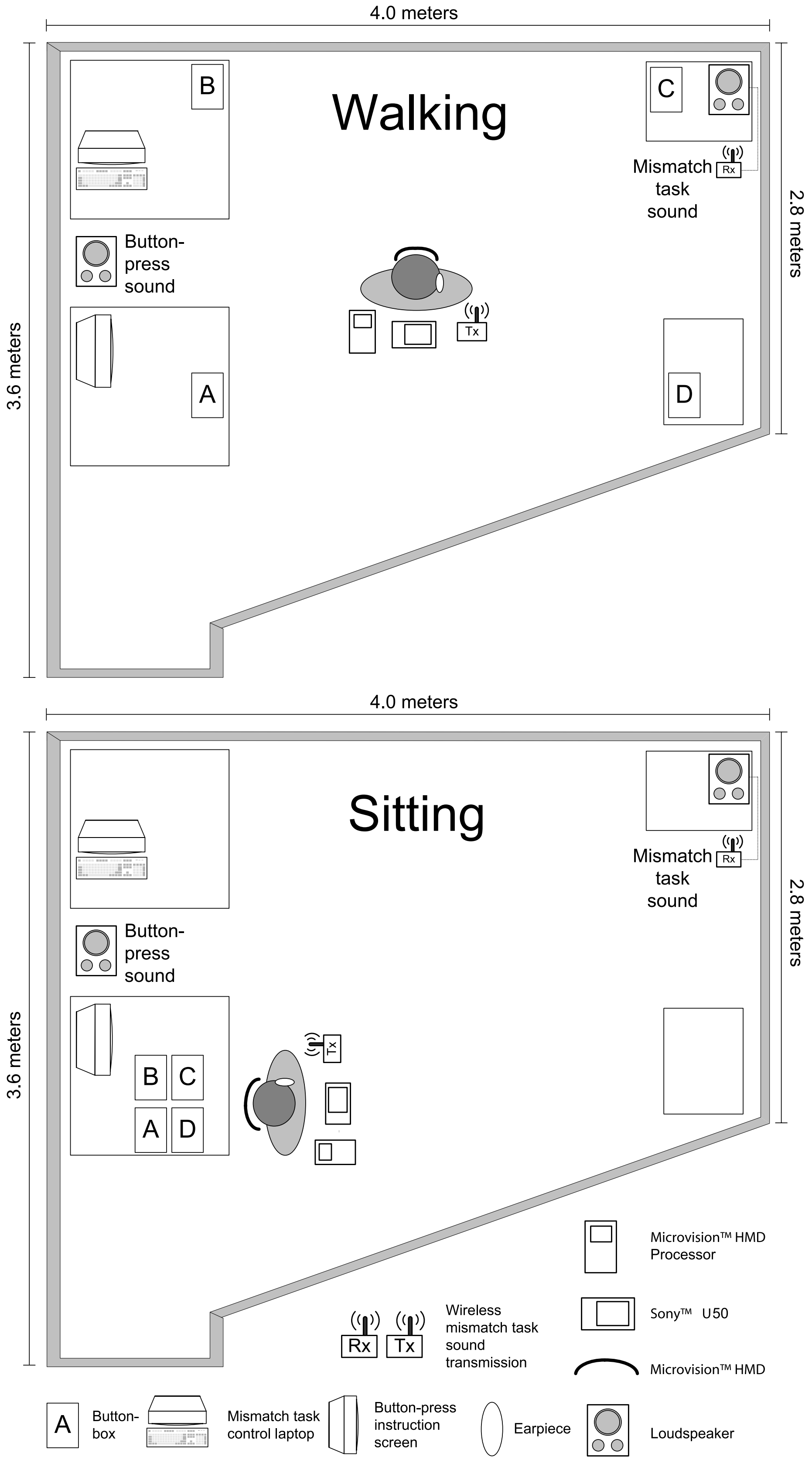

Figure 2. Experimental layouts for the walking and sitting conditions.

Figure 3. Participant in walking condition (top panel) and in sitting condition (bottom panel).

Counterbalancing. Order of presentation of the four conditions was counterbalanced in a Latin square design. Across each successive group of 4 participants, each condition was preceded and followed by all other conditions equally often, and each condition appeared at each serial position equally often. In each experimental condition there were two separate trials of 4 min each, one using a scenario with a relatively high rate of mismatches and the other using a scenario with a relatively low rate. Over the whole experiment, each experimental condition occurred in combination with all eight mismatch task scenarios and in all serial positions.

Tasks

Mismatch task. Three objects moved around a screen bouncing off each other and off the walls (see Figure 1). There were two soft objects and one hard object. The surrounding wall was defined as hard. Participants kept a silent mental count of the number of times the visual and auditory behavior of the objects mismatched. The correct matching sounds and incorrect mismatch sounds when objects collided were

- soft object hits soft object ⟶ soft sound (match)/ hard sound (mismatch);
- hard object and soft object hit ⟶ soft sound (match)/ hard sound (mismatch);
- hard object hits wall ⟶ hard sound (match)/soft sound (mismatch); and
- soft object hits wall ⟶ soft sound (match)/hard sound (mismatch).

Eight different scenarios were used for the eight 4-min experimental trials. There were between 297 and 327 bounces in each scenario. In the low mismatch rate conditions, the number of mismatches was 22, 23, 27, and 27 across scenarios, making 7.0%, 7.4%, 8.6%, and 8.6% of total bounces, respectively. In the high mismatch rate conditions, the number of mismatches was 33, 34, 35, and 37 across scenarios, making 10.3%, 10.8%, 10.9%, and 11.3% of total bounces, respectively. The intermismatch interval was always greater than 3 s.

Button-press task. The button-press task ensured that participants would move around the room in the walking condition. Four button boxes, labeled A, B, C, and D, were placed at table height in each corner of the room for the walking condition and in a similar configuration in front of the participant for the sitting condition (see Figures 2 and 3). A large letter indicating the button box to press was displayed on a computer screen at the front of the room. The letter was selected quasi-randomly from the set {A, B, C, D}. The participants' task was to press the button box corresponding to the letter displayed on the computer screen. A notification sound, from a different location and with a different acoustic profile from the bounce sound, alerted participants to a change. The letter changed every 8 s, so that 30 button presses were required per trial.

Subjective workload questionnaire. NASA Task Load Index (NASA-TLX) rating scales were administered after participants completed each of the four experimental conditions (Hart & Staveland, 1988). The procedure for ranking the subjective importance of the TLX subscales was not used (Nygren, 1991).

Apparatus

The HMD was a Microvision Nomad™ ND2500 with a single optical see-through monocle (800 × 600). 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. In the free-field condition, bounce sounds generated from the U50 were sent to a loudspeaker (Edirol™ MA-7A) using 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 70 dB(A) maximum. In the earpiece condition, ear buds (Sony™ MDR-E829V, with volume control) were connected to the U50.

The button-press task was controlled with E-Prime software on a desktop computer with a 17-inch (43.2-cm) LCD display showing the 280-point font letter. The push buttons 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 dB(A) maximum.

Procedure

The volume of the free-field sound and the earpiece were equated as follows. The participant sat on a chair in the middle of the room at approximately 180 cm from the speaker. The free-field bounce sound was the standard, and the participant adjusted the earpiece sound to match it. Then participants adjusted the focus of the HMD via the hyperopic focusing method until they felt they could view the HMD display and the wall comfortably at the same focal distance (Behar, Wiley, Levine, Rash, & Walsh, 1990).

Participants learned to do the mismatch task in several phases before experimental trials began: (a) monitor objects and sound without mismatches; (b) monitor objects and sounds for mismatches (mismatch task); (c) do mismatch task and button-press task; (d) do mismatch task and button-press task while sitting with an earpiece; (e) do mismatch task and button-press task while walking with free-field sound.

Participants were asked to face each button when they pressed it and to do the button-press task as efficiently as possible while making the mismatch task their most important task. They then proceeded to the experimental trials.

RESULTS

Mismatch Accuracy

Mismatch accuracy was the number of mismatches reported divided by the total number of mismatches presented, expressed as a percentage. On the occasional trial, participants reported more than the actual number of mismatches, leading to a score greater than 100%. Usually, however, participants reported fewer than the actual number of mismatches, suggesting that misses were generally more likely than false alarms. As a result, condition averages were always less than 100%.

A 2 × 2 × 2 ANOVA was conducted on the mismatch accuracy data, with the within-subjects factors of movement (sitting vs. walking), sound delivery (free field vs. earpiece), and mismatch rate (low vs. high). As predicted, there was a significant main effect of movement in how accurately mismatches were counted, $F(1, 19) = 8.887$, $MSE = 205.9$, $p = .008$, partial $\eta^2 = .319$ (see Figure 4). Participants counted mismatches more accurately when sitting ($M = 83\%$, $SD = 15.6\%$) than when walking ($M = 76\%$, $SD = 18.7\%$). There was no main effect of sound delivery, $F(1, 19) = 1.122$, $MSE = 159.3$, $p = .303$, partial $\eta^2 = .056$. There was a significant main effect of mismatch rate, $F(1, 19) = 78.043$, $MSE = 135.6$, $p < .001$, partial $\eta^2 = .804$. Participants counted mismatches more accurately when the mismatch rate was low ($M = 88.2\%$, $SD = 18.1\%$) than when it was high ($M = 71.9\%$, $SD = 15.8\%$).

As predicted, there was a significant two-way interaction between movement and sound delivery, $F(1, 19) = 6.449$, $MSE = 229.5$, $p = .02$, partial $\eta^2 = .253$ (see Figure 4). However, the results of planned comparisons comparing earpiece versus free field at each level of movement were not as hypothesized. Instead, when participants were sitting, they counted mismatches more accurately with sound delivered in free field ($M = 87.6\%$, $SD = 17.5\%$) than through an earpiece ($M = 79.4\%$, $SD = 16.9\%$), $p = .02$. When walking, however, participants counted mismatches equally well regardless of whether sound was delivered in free field ($M = 74.7\%$, $SD = 19.6\%$) or through the earpiece ($M = 78.7\%$, $SD = 20.0\%$), *ns*.

A Tukey HSD test was used to aid interpretation of the latter interaction and indicated that when listening to the sound in free field, participants counted mismatches more accurately when they were sitting ($M = 87.6\%$, $SD = 17.5\%$) than when they were walking ($M = 74.7\%$, $SD = 19.6\%$), $p = .006$. No other post hoc contrasts were significant.

There was a significant two-way interaction between sound delivery and mismatch rate, $F(1, 19) = 5.125$, $MSE = 80.1$, $p = .036$, partial $\eta^2 = .212$. Participants counted low mismatches particularly accurately with free-field sound. No other interactions were significant.

Button Presses

As intended, button-press accuracy was extremely high in all conditions, ranging from 99.3% to 99.8% correct, so no inferential tests could be

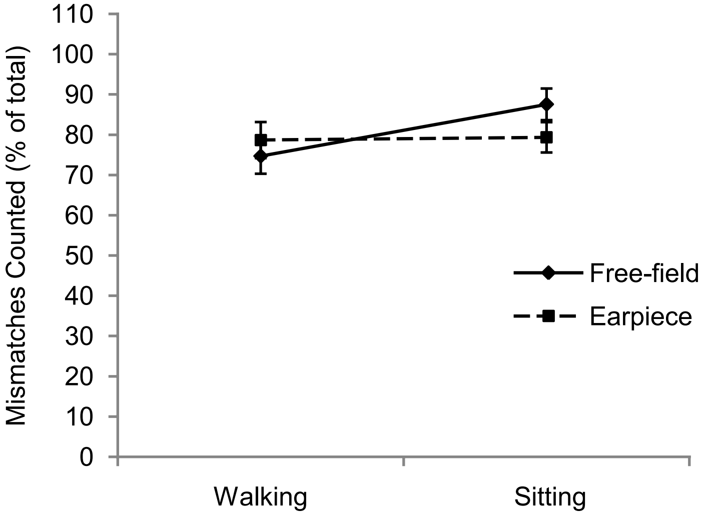

Figure 4. Mismatches counted as a percentage of actual mismatches (+ *SE*) for movement (walking vs. sitting) and sound delivery (free field vs. earpiece) conditions.

run. There was therefore no evidence of a trade-off in accuracy between tasks across the walking and sitting conditions.

A 2 (movement) × 2 (sound delivery) × 2 (mismatch rate) × 4 (button identity) repeated measures ANOVA was conducted on the button-press reaction time data. As predicted, there was a highly significant main effect of movement on reaction time, $F(1, 19) = 346.517$, $MSE = 2{,}480{,}816$, $p < .001$, partial $\eta^2 = .948$. Walking participants pressed buttons more slowly ($M = 3{,}860$ ms, $SD = 682$ ms) than sitting participants ($M = 1{,}542$ ms, $SD = 267$ ms), largely because of the time needed to walk to the button. There was also a significant main effect of button identity, $F(3, 57) = 3.116$, $MSE = 224{,}035$, $p = .033$, partial $\eta^2 = .141$. A Tukey HSD test indicated that responses to Button A were slightly faster than those to Button C, $p = .028$. No other main effects or interactions were significant.

NASA-TLX

A 2 (movement) × 2 (sound delivery) × 6 (TLX subscale) repeated measures ANOVA was run on responses to the NASA-TLX. There was a significant main effect of movement, $F(1, 18) = 45.805$, $MSE = 11.450$, $p < .001$, partial $\eta^2 = .718$. Participants rated their workload higher for walking than for sitting (see Figure 5). There also was a significant main effect of TLX subscale, $F(5, 90) = 16.123$, $MSE = 39.35$, $p < .001$, partial $\eta^2 = .473$, with mental demand rated highest and physical demand lowest in workload intensity.

A significant interaction was found between movement and TLX subscale, $F(5, 90) = 2.769$, $MSE = 7.78$, $p = .023$, partial $\eta^2 = .133$, because physical demand was disproportionately higher when participants were walking. No other main effects or interactions were significant.

DISCUSSION

The aim of this study was to assess the role of both self-motion and sound delivery method on people's ability to perform multisensory integration with an HMD. The first hypothesis was that participants would count mismatches less accurately when they were walking than when they were sitting, regardless of sound delivery method. At face value this hypothesis was supported, especially as participants also reported higher workload when walking. However, these results must be considered in the context of the interaction between movement and sound delivery.

The second hypothesis was that when participants were sitting they would count mismatches equally well with earpiece and free-field sound but that when walking they would count mismatches less accurately with free-field sound than with an earpiece. This hypothesis was not supported. First, performance with the earpiece was less effective overall than predicted. When participants were

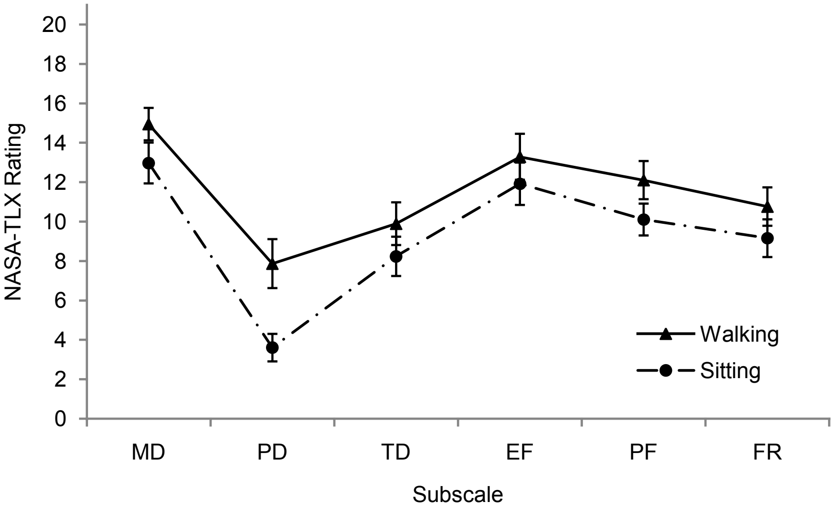

Figure 5. NASA-TLX subjective workload ratings (±*SE*) for walking and sitting conditions. Abbreviations for subscales are MD = mental demand, PD = physical demand, TD = temporal demand, EF = effort, PF = performance, and FR = frustration.

seated they counted mismatches more accurately with free-field sound than with the earpiece. Second, when sound came from the earpiece, participants counted mismatches equally well whether they were sitting or walking (79% of total mismatches), a finding inconsistent with the first hypothesis and with Sparrow et al. (2002) and Lajoie et al. (1993). When sound came from free field, however, participants counted fewer mismatches when walking (74% of total mismatches) than when sitting (87% of total mismatches).

The worse performance when participants were walking and listening to free-field sound cannot be due to walking alone, as walking did not hurt performance when participants used an earpiece. Similarly, the result cannot be due to free-field sound alone, as free-field sound did not hurt performance when participants were seated.

Instead, when walking participants heard the sounds in free field, they may have counted fewer mismatches than when they were seated because of the combination of three factors.

1. the rapid arrival rate of bounces, requiring quick decisions as to whether or not a mismatch had happened;
2. the need to time-share watching the visual objects with locating and acquiring the button boxes; and
3. the inconsistent spatial and motion relationships between visual events and free-field sound.

The first two factors were shared with other conditions, but they may have exacerbated the effect of the critical third factor.

First, the relatively rapid arrival rate of bounces and the low a priori probability of mismatches probably made participants less likely to increment their mismatch count when they were not quite sure whether a mismatch had occurred.

Second, participants could not fixate all bounces, especially while also locating and acquiring the button boxes. When an object moves across the visual field outside foveal vision, participants often misjudge the timing and location of the object (Brenner, van Beers, Rotman, & Smeets, 2006). Bounces seen outside foveal vision may be subject to such misjudgments and may have been captured by the wrong bounce sound (Alais & Burr, 2004; Ernst & Banks, 2002; Lewald, Ehrenstein, & Guski, 2001).

Third, and most critically, the inconsistent spatial and motion relationships between vision and free-field sound for walking participants may occasionally have biased the order in which events seemed to occur or may have caused the wrong sound to be paired with the wrong visual event. Such biases and mispairings have been frequently reported in the multisensory integration literature (Fendrich & Corballis, 2001; Jaekl & Harris, 2007; Morein-Zamir, Soto-Faraco, & Kingstone, 2003; Soto-Faraco et al., 2006). In the mismatch task, 14% of mismatches occurred sufficiently close to neighboring matches (within around 200 ms) that spurious matches were possible under ambiguous conditions. This figure is consistent with the 13% difference we found in mismatch count accuracy between the free-field walking and sitting conditions. This speculation should be pursued in future basic and applied studies because it suggests mechanisms by which self-motion may affect multisensory integration.

A potential concern is that the present findings may be specific to the situation in which participants maintain a count of mismatches in running memory. However, it is unclear why the working memory load of maintaining a running count should interfere selectively with the walking free-field condition to create the interaction shown in Figure 4. If the workload of maintaining a running memory count had interacted with the workload of walking, for example, then an equal decrement in both walking conditions would have been found, but it was not.

Limitations and Future Research

This study has several limitations, all of which indicate directions for future research. First, our current dependent measure does not let us probe the specific conditions that made mismatches hard to detect, such as when participants were walking and listening to the sound in free field. In future research, participants could indicate mismatches with a button click so that the exact time of the button click, relative to bounces, could be captured. Hits, false alarms, misses, and correct rejections could be discriminated. Such an analysis would indicate, in each condition, consistencies in the arrangement of stimuli that lead to false alarms or misses. However, it may be difficult during analysis to associate a click with the appropriate bounce. Moreover, preliminary investigations suggest that requiring participants to use a button click may increase perceptual-motor load and may sometimes even increase the overall load on working memory because it is harder to compensate for inadvertent button clicks than for a mental miscount.

Second, only a task requiring multisensory integration of sound and vision has been considered. The effect of sound delivery method and self-motion on tasks that can be performed with vision or sound alone, or in which sound is ancillary to vision, is still unknown, yet such tasks may represent more common uses of HMDs.

Third, the mismatch task involves a dynamic display on the HMD. Whether a task using a static display (e.g., Laramee & Ware, 2002) would be sensitive to sound delivery method and self-motion needs to be tested.

Fourth, overall performance with the earpiece was not as good as expected. A higher quality earpiece or a better match to free-field sound dynamic range may improve performance. Moreover, headphones would place the perceived location of the sound in the center of the head, which may increase spatial congruence and improve performance.

CONCLUSIONS

Advanced displays offer new ways of accessing information, but they can also break otherwise invariant relationships between information from different modalities (Gibson, 1979). When displayed on an HMD, visual information moves with the head, but unless complex and expensive spatialization is used, sounds will not be perfectly mapped to the location of associated visual objects. Our results indicate that multisensory integration performance suffers particularly badly if sound is played in free field while an HMD wearer walks about, probably because the relationship between vision and sound changes continuously. Further testing will establish the generalizability of these results, but the lack of any prior information in the literature may reflect shortcomings in how advanced displays are evaluated.

Lightweight HMDs intended for use by mobile users in industrial, military, or medical contexts must be effective regardless of the user's location, orientation, and motion (Liu et al., 2008; Sanderson et al., 2008). Most prior research on motion and auditory stimuli investigates a stationary participant's perception of moving auditory stimuli, rather than a moving participant's perception of head-referenced versus world-referenced auditory stimuli.

There is surprisingly little research on factors influencing people's ability to use information in visual and auditory modalities while they walk about, yet it is a common human task increasingly supported by personal technologies such as personal digital assistants, iPods™, and HMDs. New display technologies underscore the need for research that extends current theories of multisensory information processing (Calvert, Spence, & Stein, 2004) and ecological perception (Gibson, 1979; Neuhoff, 2004) to the important problem of the mobile user.

ACKNOWLEDGMENTS

We thank Stas Krupenia and David Liu for their pioneering work with head-mounted displays within our laboratory and for their help; members of the Cognitive Engineering Research Group for comments on the manuscript; Jason Mattingley and David Alais for helpful discussions; and Phil Cole, Nick Sebald, and Mark Corben for technical assistance. This research was supported by strategic funding from the School of Psychology at The University of Queensland and by Australian Research Council Discovery Grant DP0559504. We thank three anonymous reviewers whose comments greatly improved the paper.

REFERENCES

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14,* 257–262.

Barrass, S., & Kramer, G. (1999). Using sonification. *Multimedia Systems, 7,* 23–31.

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.

Bolia, R. S., D'Angelo, W. R., & McKinley, R. L. (1999). Aurally aided visual search in three-dimensional space. *Human Factors, 41,* 664–669.

Brenner, E., van Beers, R. J., Rotman, G., & Smeets, J. B. J. (2006). The role of uncertainty in the systematic spatial mislocalization of moving objects. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 811–825.

Brewster, S. A. (1994). Providing a structured method for integrating non-speech audio into human-computer interfaces (Doctoral dissertation, University of York, 1990). *Dissertation Abstracts International, 57,* 1340.

Calvert, G. A., Spence, C., & Stein, B. E. (2004). *The handbook of multisensory processes.* Cambridge, MA: MIT Press.

Driver, J., & Spence, C. (1994). Spatial synergies between auditory and visual attention. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance XV* (pp. 311–331). Cambridge, MA: MIT Press.

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

Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception and Psychophysics, 63,* 719–725.

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience, 11,* 473–490.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Guski, R., & Troje, N. F. (2003). Audiovisual phenomenal causality. *Perception and Psychophysics, 65,* 789–800.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdam: North-Holland.

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.

Jaekl, P. M., & Harris, L. R. (2007). Auditory-visual temporal integration measured by shifts in perceived temporal location. *Neuroscience Letters, 417,* 219–224.

Kallinen, K., & Ravaja, N. (2007). Comparing speakers versus headphones in listening to news from a computer – Individual differences and psychophysiological responses. *Computers in Human Behavior, 23,* 303–317.

Kramer, G. (1994). An introduction to auditory display. In G. Kramer (Ed.), *Auditory display: Sonification, audification, and auditory interfaces* (pp. 1–78). Reading, MA: Addison Wesley.

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.

Lewald, J., Ehrenstein, W. H., & Guski, R. (2001). Spatio-temporal constraints for auditory-visual integration. *Behavioral Brain Research, 121,* 69–79.

Lewald, J., & Guski, R. (2004). Auditory-visual temporal integration as a function of distance: No compensation for sound-transmission time in human perception. *Neuroscience Letters, 357,* 119–122.

Liu, D., Jenkins, S., Sanderson, P. M., Leane, T., Watson, M. O., & Russell, W. J. (2008). Simulator evaluation of head-mounted displays for patient monitoring. *Anesthesia and Analgesia, 106*(Suppl. 4), S34.

Michotte, A. (1963). *The perception of causality* (T. Miles & E. Miles, Trans.). New York: Basic Books. (Original work published 1954)

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research, 17,* 154–163.

Nelson, T. M., & Nilsson, T. H. (1990). Comparing headphone and speaker effects on simulated driving. *Accident Analysis and Prevention, 22,* 523–529.

Neuhoff, J. G. (Ed.). (2004). *Ecological psychoacoustics.* Amsterdam: Elsevier Academic Press.

Newell, F. N. (2004). Cross-modal object recognition. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 123–138). Cambridge, MA: MIT Press.

Nygren, T. E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors, 33,* 17–33.

Ockerman, J., & Pritchett, A. (1998). Preliminary investigation of wearable computers for task guidance in aircraft inspection. In *Proceedings of the International Symposium on Wearable Computers (ISWC '98)* (pp. 33–40). Pittsburgh, PA: CS Press.

Ong, S. K., Yuan, M. L., & Nee, A. Y. C. (2008). Augmented reality applications in manufacturing: A survey. *International Journal of Production Research, 46,* 2707–2742.

Ormerod, D. F., Ross, B., & Naluai-Cecchini, A. (2003). Use of an augmented reality display of patient monitoring data to enhance anesthesiologists' response to abnormal clinical events. *Studies in Health Technology and Informatics, 94,* 248–250.

Patterson, R, Winterbottom, M. D., & Pierce, B. J. (2006). Perceptual issues in the use of head-mounted visual displays. *Human Factors, 48,* 555–573.

Perrott, D. R., Cisneros, J., McKinley, R. L., & D'Angelo, W. R. (1996). Aurally aided visual search under virtual and free-field listening conditions. *Human Factors, 38,* 702–715.

Radeau, M. (1994). Auditory-visual spatial interaction and modularity. *Current Psychology of Cognition, 13,* 3–51.

Roibas, A. C., Geerts, D., Furtado, E., & Calvi, L. (2008). Implications of the socio-physical contexts when interacting with mobile media. *Personal and Ubiquitous Computing, 12,* 279–280.

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

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

Soto-Faraco, S., Kingstone, A., & Spence, C. (2006). Integrating motion information across sensory modalities: The role of top-down factors. *Progress in Brain Research, 155,* 273–286.

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.

Spence, C., & Driver, J. (2004). *Crossmodal space and crossmodal attention.* New York: Oxford University Press.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26,* 212–215.

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

Viaud-Delmon, I., Warusfel, O., Seguelas, A., Rio, E., & Jouvent, R. (2006). High sensitivity to multisensory conflicts in agoraphobia exhibited by virtual reality. *European Psychiatry, 21,* 501–508.

Vroomen, J., & Keetels, M. (2006). The spatial constraint in intersensory pairing: No role in temporal ventriloquism. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 1063–1071.

Walker, B. N., & Kramer, G. (2004). Ecological psychoacoustics and auditory displays: Hearing, grouping, and meaning making. In J. G. Neuhoff (Ed.), *Ecological psychoacoustics* (pp. 150–174). Amsterdam: Elsevier Academic Press.

Watson, M. O., & Sanderson, P. (2004). Sonification supports eyes-free respiratory monitoring and task time-sharing. *Human Factors, 46,* 497–518.

Watson, M. O., & Sanderson, P. (2007). Designing for attention with sound: Challenges and extensions to ecological interface design. *Human Factors, 49,* 331–346.

Yeh, M., Merlo, J. L., Wickens, C. D., & Brandenburg, D. L. (2003). Head up versus head down: The costs of imprecision, unreliability, and visual clutter on cue effectiveness for display signaling. *Human Factors, 45,* 390–408.

Matthew Thompson is the research coordinator in the Cognitive Engineering Research Group (CERG) at The University of Queensland. He received a dual BSc(Hons) in psychology and BInfTech in 2007 from The University of Queensland.

Penelope Sanderson is a professor of cognitive engineering and human factors at The University of Queensland. She received her Ph.D. in engineering psychology in 1985 from the University of Toronto.

Date received: February 12, 2008
Date accepted: June 9, 2008