Multisensory Integration With a Head-Mounted Display: Role of Mental and Manual Load

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

Objective: The aim of this study was to replicate the finding that multisensory integration with a head-mounted display (HMD) is particularly difficult when a person is walking and hearing sound from a free-field speaker, and to extend the finding with a response method intended to reduce workload. **Background:** HMDs can support the information needs of workers whose work requires mobility, but some low-cost solutions for delivering auditory information may be less effective than others. **Method:** For the study, 24 participants detected whether shapes moving on the HMD screen made a sound appropriate to their forms when they collided with other shapes. Independent variables were self-motion (participants were mobile or seated), sound delivery (free-field speakers or an earpiece), and response method (noting mismatches via a mental count or via a manual clicker). **Results:** Unexpectedly, overall mismatch task accuracy was worse with the clicker (p = .027) than without. Participants also reported that it was harder to time-share the mismatch task with clicker responses (p = .033). In the clicker condition, self-motion and sound delivery interacted but in the opposite direction to the previous study. **Conclusion:** The best way of delivering auditory information to mobile workers performing a multisensory integration task with an HMD may depend on whether responding involves mental load or manual load. Broader theories are needed to capture factors influencing performance. **Application:** Until more powerful theory is developed, designers should perform careful formative and summative tests of whether the activities to be performed by mobile HMD wearers will make some sound delivery solutions less effective than others.

INTRODUCTION

In the natural world, events and objects that we perceive with more than one sense are usually experienced as integrated perceptual wholes because the sensory information usually originates from the same time and same point in space (Bertelson & de Gelder, 2004; Newell, 2004). In contrast, advanced displays designed to support mobile workers, such as head-mounted displays (HMDs), auditory displays, and tactile displays, may present information about a virtual event or object in different modalities that are also separated in time and space, so breaking otherwise invariant relationships.

Unfortunately, little is known about how effectively people integrate multimodal information when using an HMD and moving around the environment. Multisensory integration is usually explored in tightly controlled laboratory experiments with the participant seated at a computer. Applied work involving multisensory integration has often focused on work situations in which people are seated, such as aviation and driving (Ho & Spence, 2008). Important previous research on HMDs themselves has also been conducted with seated participants (Laramee & Ware, 2002). It is unclear what we might find with participants who are mobile.

In this study, we examine participants' ability to perform multisensory integration with an HMD using different methods of delivery of sound while participants are either mobile or seated. Specifically, we explore two ways that

Address correspondence to Matthew B. Thompson, School of Psychology, The University of Queensland, St. Lucia QLD, Australia 4072; mbthompson@gmail.com. ***HUMAN FACTORS,*** Vol. 52, No. 1, February 2010, pp. 92–104. DOI: 10.1177/0018720810367914.

participants may provide responses, which lets us explore forms of responding that are representative of wearable computing tasks and test the generality of prior research (Harrison, Thompson & Sanderson, 2010 [this issue]; Thompson & Sanderson, 2008). Comparing two forms of response also lets us test the possible impact of working memory load on performance.

Advanced Displays in Safety-Critical Systems

Our program of research was motivated by the practical issue of how to present visual and auditory information about a patient to anesthesiologists in the operating theater, but it extends to other contexts. HMDs and advanced auditory displays have been evaluated in patient care settings to determine whether they help anesthesiologists monitor patients (Liu, Jenkins, & Sanderson, 2009; Liu, Jenkins, Sanderson, Watson, et al., 2009; Sanderson et al., 2008).

Although HMDs appear to confer some benefits, it is unclear how best to deliver sound when visual information from an HMD is supplemented with auditory displays. For example, aspects of waveform-based information can also be sonified, as currently exists when electrocardiogram or plethysmography signals are fed to the "beeping" heart monitor and as may exist in the future with capnography waveforms and respiratory sonification (Sanderson et al., 2008; Watson & Sanderson, 2004, 2007). Sonification (continuous auditory display) offers the anesthesiologist good peripheral awareness of a patient's status while focal attention is elsewhere, whereas visual displays support focal attention.

As noted earlier, people perceive objects and events containing information from multiple modalities better when the information is presented at the same time and in the same location. For example, if visual and auditory displays of patient vital signs are integrated temporally and spatially, they may improve anesthesiologists' ability to extract information from each display, to transition between attentional modes, and to detect physiological or equipment-based anomalies. Factors making spatial and temporal integration difficult may compromise these abilities.

Moreover, without spatial integration, the full potential of displays may not be realized. It has been noted that anesthesiologists who can see all vital signs on an HMD often respond to an alarm coming from free-field speakers on the anesthesia workstation by unnecessarily turning to the workstation (D. Liu, personal communication, September 30, 2009). A sound source closer to the HMD display that orients attention to the HMD and supports the perceptual integration of sound and vision could be much more effective.

However, it is hard to engineer the visual and auditory information for an event so that both appear to come from the exact same point in space (Gray, Tan, & Young, 2002). Techniques such as head-related transfer functions can make sounds appear as if they are coming from the "virtual" location of visual information on the HMD (Wenzel, Arruda, Kistler, & Wightman, 1993), but such techniques are complex and expensive. It is unrealistic to suggest that they should be used in work domains such as anesthesia (Georgiou & Kyriakakis, 1999; So, Leung, Braasch, & Leung, 2006).

Alternatively, auditory information can be delivered publicly via a free-field speaker in a fixed position or privately via an earpiece that moves with the wearer. Recent research suggests that an earpiece may be better because it is located in close peripersonal space (Ho & Spence, 2009). In addition, we have found that an earpiece does not produce a decrement in performance when the HMD user is mobile, whereas free-field delivery of sound does produce such a decrement (Harrison et al., 2010 [this issue]; Thompson & Sanderson, 2008).

Barriers to Multisensory Integration With Advanced Displays

As indicated earlier, there is a gap in the literature on the effect of self-motion and sound delivery method on people's ability to perform any form of multisensory integration, let alone with an HMD. Walking creates background motion and is known to have intrinsic workload (Lajoie, Teasdale, Bard, & Fleury, 1993), and different methods of sound delivery have different acoustic properties that may interact with walking.

To investigate these issues, Thompson and Sanderson (2008) created a multisensory integration task that participants performed while

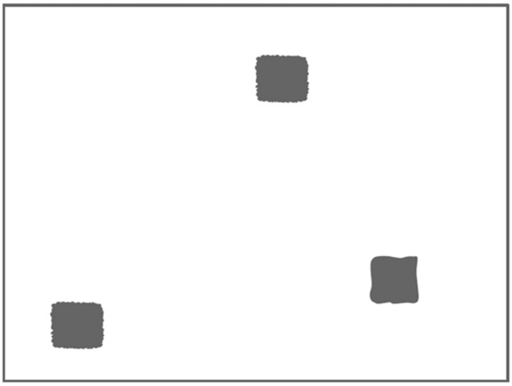

Figure 1. Shapes on the head-mounted display for the mismatch task. The top central shape and lower left shape are soft and make a soft sound when they collide with any other shape. The lower right shape and the wall are hard and make a hard sound only when the hard shape collides with the wall.

Source. Adapted 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. Adapted with permission.

either walking or sitting and while using either an earpiece or a free-field speaker: the "mismatch" task. The mismatch task was based on the well-established Michotte (1954/1963) launch task. It was intended to provide a sensitive first test of the impact of sound delivery and mobility on HMD users while retaining some properties representative of work domains in which advanced visual and auditory displays and wearable computers are used; the integration of waveforms and auditory displays in anesthesia previously mentioned is an example. Conditions in which multisensory integration errors appear may also be conditions in which various other multimodal information integration tasks become more difficult.

In the mismatch task, participants viewed two soft shapes and one hard shape moving within a rectangular wall (see Figure 1) on the HMD. The shapes bounced off each other and off the walls, making a sound with every bounce. The bounce sound was usually consistent with the visual properties of the shapes involved in the bounce on the HMD (i.e., soft vs. hard), but occasionally, participants would hear a bounce sound that was inconsistent—a mismatch. Participants kept a silent mental count of the number of times the vision and sound of a bounce mismatched.

When participants performed the mismatch task while listening to the sound through an earpiece, their performance was the same whether they were walking or sitting. When participants performed the mismatch task while listening to the sound from a free-field speaker, however, their performance was significantly worse when they were walking than when sitting.

Thompson and Sanderson (2008) speculated that the findings may be attributable to the highly dynamic spatial inconsistencies between the locations and motion of the shapes on the HMD and the fixed-source location of the free-field sound, which is constantly changing relative to the position of the moving user. A subsequent experiment by Harrison et al. (2010 [this issue]) with free-field sound replicated the results and suggested a strong role of background motion as well as relative sound location. Overall, Thompson and Sanderson's results are difficult to explain fully with theories of multisensory integration, sound source location, or background motion.

Memory Load and Response Modality

A potential limitation of the Thompson and Sanderson (2008) and Harrison et al. (2010 [this issue]) procedure noted by several earlier reviewers is that participants were asked to keep a running mental count of mismatches and to report the total at the end of each trial. A concern was that the memory load associated with keeping a mental count could somehow have caused the pattern of results. Carlson and Cassenti (2004) have shown that keeping a mental count imposes high mental workload and is prone to various errors. In the mismatch task, errors could be *judgment errors* (false alarms, misses, and perceptual misses of the mismatch event itself) or *increment errors* (false increments of the current total in working memory or false nonincrements).

Although working memory is undoubtedly involved in the mismatch task, it is probably not solely responsible for the unique performance decrement seen when participants are walking with free-field sound (Thompson & Sanderson, 2008). Working memory is required equally across all conditions of the mismatch task, and it would have to interact with self-motion and free-field sound delivery to explain Thompson and Sanderson's (2008) results.

However, concerns about the mental count measure are common, and they have practical as well as methodological implications. When using wearable computing, people respond vocally and physically as well as mentally. Therefore, the generality of the Thompson and Sanderson (2008) findings to other forms of responding needs to be explored. For example, participants might register mismatches by either (a) using a clicker or other manual response or (b) making a vocal response, both of which would remove the requirement to store and rehearse the current total. Either kind of response would be less mentally loading than the mental count and would be a more precise measure of mismatch detection that lets us explore the impact of sound delivery and mobility on accuracy, sensitivity, and response bias.

Compared with the mental count, the clicker may introduce greater motor demand, whereas a vocal response may interfere with the participant's ability to hear the auditory signals during the experiment. Because interference with the auditory signals would be a direct experimental confound, we decided to compare responding via the clicker with the mental count procedure to see whether the Thompson and Sanderson (2008) results generalize.

Hypotheses

The first hypothesis (Hypothesis 1) is that compared with keeping a mental count, registering mismatches with a clicker should improve mismatch task accuracy across all conditions. A clicker should reduce participants' reliance on working memory and remove errors associated with demands on, or failures of, working memory (Carlson & Cassenti, 2004).

The second hypothesis (Hypothesis 2) is that when participants keep a mental count of mismatches, Thompson and Sanderson's (2008) results will be replicated. When participants wear an earpiece, their mismatch task accuracy will be the same regardless of whether they are walking or sitting. When participants hear the sounds in free field, however, their mismatch task accuracy will be worse when they are walking than when they are sitting.

The third hypothesis (Hypothesis 3) is that if working memory interacts with whether the participant is walking and how sound is being delivered, then with the clicker there should be no interaction between whether the participant is walking and how sound is delivered. In contrast, if the interaction seen in Hypothesis 2 remains with the clicker, then the interaction cannot have been caused by working memory. A separate signal detection analysis on conditions in which the clicker is used should also show no interaction.

METHOD

Participants

The study was approved by The University of Queensland's Behavioural and Social Sciences Ethical Review Committee. Participants were 12 female and 12 male ($N = 24$) university students between 18 and 25 years of age ($M = 21.88$, $SD = 1.90$) who earned course credit or cash for participation. All gave written informed consent and reported normal or corrected-to-normal vision and normal hearing.

Design

There were three variables in this experiment, each with two levels, for a total of eight conditions administered on a within-subjects basis: Self-motion (walking vs. sitting) × Sound Delivery (free field vs. earpiece) × Response Method (mental count vs. clicker). In contrast with Thompson and Sanderson (2008), only scenarios with a "high" rate of mismatches were used in the current experiment, because previous results showed that scenarios with a high rate of mismatches were more sensitive to changes in self-motion and sound delivery.

Each scenario lasted for 4 min. There were between 315 and 344 bounces in each scenario, and approximately 11% of the bounces were mismatches. To help control the difficulty of the task, the time interval between bounces was constrained so that no more than 30 bounces occurred within each of 0 to 99 ms, 100 to 199 ms, or 200 to 299 ms of another bounce. The average interbounce interval was 709 ms, ranging from 0 ms to 3,963 ms, and the average intermismatch interval was 6,696 ms, ranging from 3,004 ms to 18,967 ms.

Sound delivery. Sound was delivered either in free field or via earpiece. In the free-field condition, the bounce sounds came from a speaker in the corner of the room. In the earpiece condition, the bounce sounds came from an earpiece in the participant's right ear.

Self-motion. Self-motion was manipulated by participants either walking or sitting. In the walking condition, four button boxes were placed in each corner of the room so participants had to walk to them (see Figure 2). In the sitting condition, the participant sat on a chair at a desk with their head stabilized on a chin rest.

Response method. Participants responded by either reporting a mental count or by pressing a clicker. In the mental count condition, participants kept a silent mental count of the number of mismatches they detected. In the clicker condition, participants indicated that they saw a mismatch by pressing a clicker. Participants were instructed to hold the clicker in their right hand and to use their left hand to press button boxes.

Counterbalancing. To avoid sequencing or carryover effects, the order of presentation of the eight conditions was counterbalanced with the use of a Latin square. Across each successive group of 8 participants, each condition was preceded and followed by all other conditions equally often, and each condition appeared at each serial position equally often. Each experimental condition was observed with all mismatch bounce scenarios and in all serial positions.

Apparatus

The HMD was a Microvision Nomad™ ND2000 with a single optical see-through monocle (800 × 600 pixel display). The mismatch task shapes displayed on the HMD were received from a Sony™ U50 tablet computer, which ran the mismatch task. The mismatch task shapes were 80 × 80 pixels and moved at 150 pixels/s. In the free-field condition, the bounce sounds were sent to a loudspeaker via a wireless transmitter and receiver. The sound pressure level (SPL) of the sounds, measured from the center of the room, was 70 dBA on the max hold reading which captured the highest reading. In the earpiece condition, earbuds (Sony MDR-E829V, with volume control) were connected through the wireless sound transmitter and receiver to control for any differences in transmitting sound wirelessly compared with wired.

The button-press task was controlled with E-Prime software (Psychology Software Tools, Pittsburgh, PA) on a desktop computer with a 17-in. 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 speaker was approximately 64 dBA on the max hold reading.

Tasks and Measures

Mismatch task. Three shapes moved around a screen bouncing off each other and off the walls (see Figure 1). There were two soft shapes and one hard shape. 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 shapes mismatched. The correct matching sounds and incorrect mismatch sounds when shapes collided were as follows:

- Soft shape hits soft shape: soft sound = match; hard sound = mismatch
- Hard shape and soft shape hit: soft sound = match; hard sound = mismatch
- Hard shape hits wall: hard sound = match; soft sound = mismatch
- Soft shape hits wall: soft sound = match; hard sound = mismatch

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 in the walking condition and in a similar configuration in front of the participant in the sitting condition (see Figure 2). 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 quasirandomly 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.

Questionnaires. After each of the eight conditions, participants responded to questionnaires about how easy it was (a) to detect mismatches,

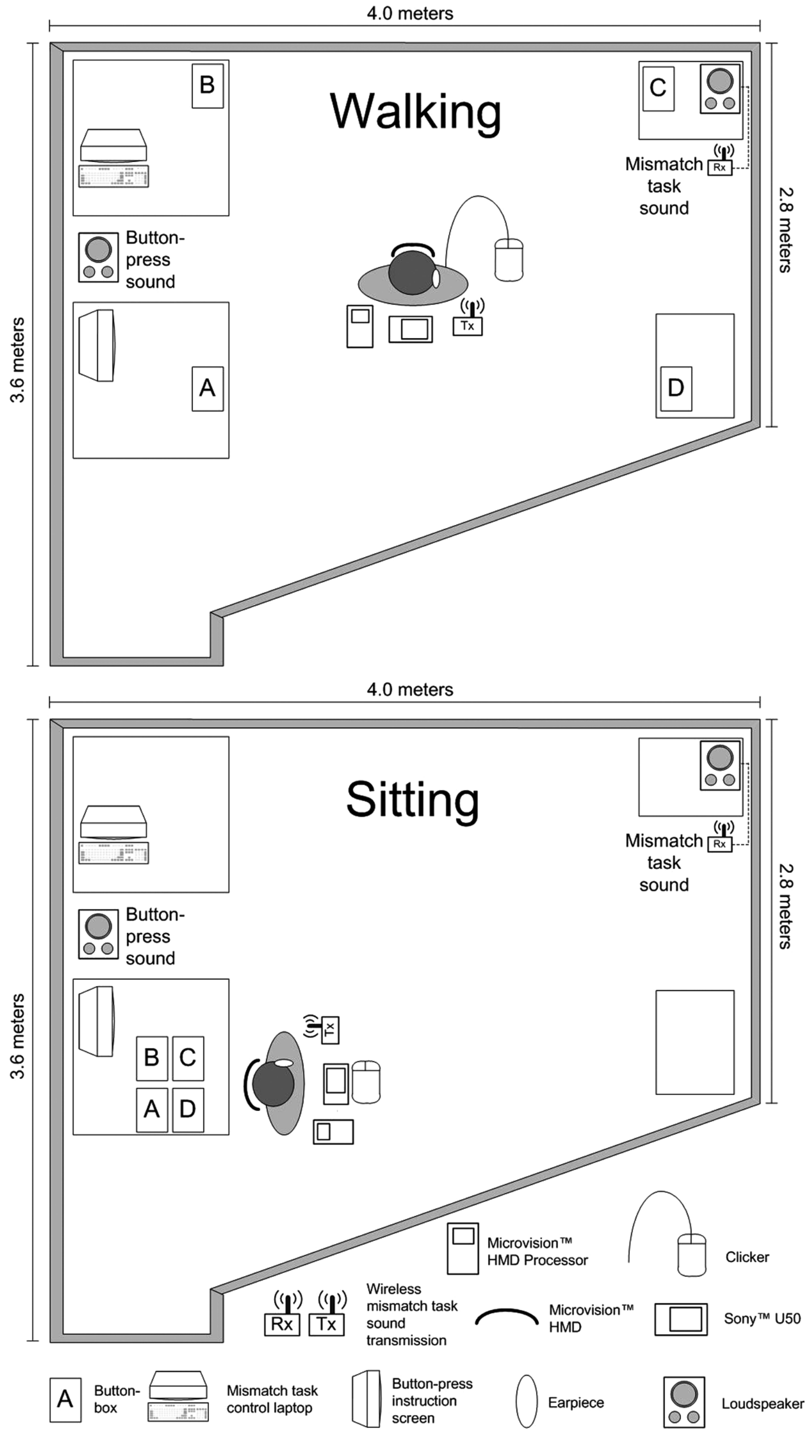

Figure 2. Experimental setup for sitting and walking conditions. Buttons are arranged in the sitting condition so that they match how they were arranged in the walking condition. When participants were not using the clicker, it was attached to their belt, as demonstrated in the sitting condition. The clicker was used in both the sitting and walking conditions; it is shown here in just the walking condition for purposes of illustration.

(b) to react to mismatches, (c) to integrate visual information with sounds, and (d) to time-share monitoring for mismatches with pushing the buttons. They also completed the National Aeronautics and Space Administration–Task Load Index (NASA-TLX) subjective workload rating scale (Hart & Staveland, 1988) without the subscale ranking procedure (Nygren, 1991).

Procedure

Participants read an information sheet and signed a consent form. They were then trained and given practice in the following sequence: (a) Monitor objects and sound without mismatches; (b) monitor objects and sound for mismatches (mismatch task); (c) using the method of constant stimuli, calibrate the volume of the earpiece so that it is the same volume as the free-field speakers; (d) fit and adjust HMD focus so visual stimuli and surrounding walls of the room can be viewed as same focal distance (Behar, Wiley, Levine, Rash, & Walsh, 1990); (e) do mismatch task and button-press task simultaneously; (f) practice using clicker (they were instructed to not count in head and to not add or withhold a click if they believed they had incorrectly missed or registered a mismatch); (g) undertake final full practice runs with mental count and clicker. The total number of mismatches varied across scenarios, and participants were not given any feedback about their accuracy.

RESULTS

Mismatch Accuracy Data

Mismatch accuracy data were analyzed in two ways. First, the mental count reported or the number of clicks registered by the participant was divided by the actual number of mismatches for the scenario and was reported as a percentage. Second, for the clicker condition, a signal detection analysis was performed. Responses occurring less than 200 ms after or more than 1,500 ms after a mismatch were considered false alarms. If participants responded between 200 and 1,500 ms after a mismatch, their response was deemed to be in response to the mismatch and therefore a hit; if there was no response, it was a miss. If additional bounces occurred within 1,500 ms after a mismatch, they were ignored for purposes of determining how the participant responded to the mismatch.

The lower bound of 200 ms conservatively contains the time necessary for a simple reaction (see Welford, 1980, for a summary). The upper bound of 1,500 ms provides a reasonable window of opportunity for the participants to respond, given the complexity of the task and the fact that they were asked to respond without delay rather than as quickly as possible. Adjustments to the cutoff times did not significantly change the pattern of results.

Mismatch Count Accuracy

A 2 × 2 × 2 repeated-measures ANOVA was conducted on the mismatch count data with the within-subjects factors of response method (mental count vs. clicker), self-motion (walking vs. sitting), and sound delivery (earpiece vs. free field). Means are shown graphically in Figure 3.

A significant main effect of response method was observed, $F(1, 23) = 5.548$, $MSE = .025$, $p = .027$. Contrary to the first hypothesis, when participants maintained a silent mental count, they counted mismatches more accurately (M = 77%, SD = 16%) than when they used the clicker (M = 72%, SD = 14%) rather than less. There was no significant main effect on mismatch count accuracy of either self-motion ($p = .288$) or sound delivery ($p = .794$).

The second hypothesis was that within the mental count condition, there would be a two-way interaction between self-motion and sound delivery, replicating Thompson and Sanderson (2008). A separate 2 × 2 repeated-measures ANOVA was performed on the mental count data. The two-way interaction showed a trend in the opposite direction from that predicted, but it was not significant, $F(1, 23) = 2.974$, $MSE = .010$, $p = .098$. Neither of the main effects was significant.

The third hypothesis was that removing working memory with the clicker would remove the interaction between self-motion and sound delivery in the clicker condition. A separate 2 × 2 repeated-measures ANOVA was performed on the clicker data. Unexpectedly, the interaction was significant, $F(1, 23) = 7.607$, $MSE = .007$, $p = .011$, and it was in the opposite direction from that predicted on the basis of Thompson

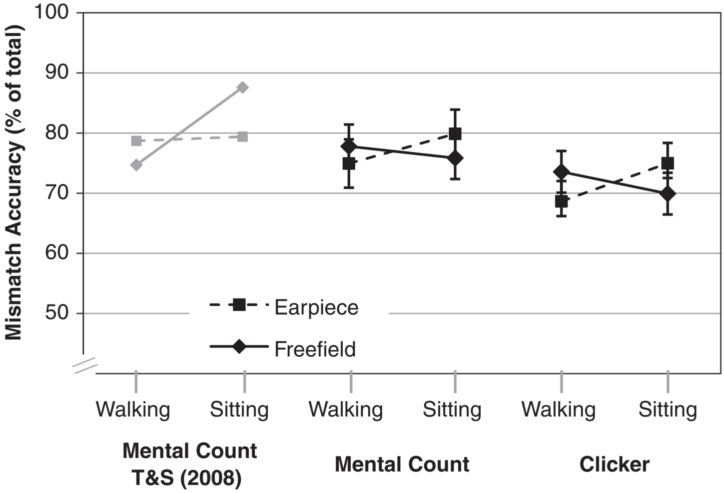

Figure 3. Mismatch accuracy, calculated as a percentage of actual mismatches (+/– standard error of the mean) for movement (walking vs. sitting) and sound delivery (freefield vs. earpiece) conditions. Results for Thompson and Sanderson (2008) are shown in the leftmost panel. For the present experiment, mental count results are shown in the center panel, and clicker results are shown in the rightmost panel.

and Sanderson (2008). When using the earpiece, participants performed slightly worse when walking ($M = 68\%$, $SD = 17\%$) than when sitting ($M = 75\%$, $SD = 12\%$), whereas when using free-field sound, participants performed slightly worse when sitting ($M = 70\%$, $SD = 17\%$) than when walking ($M = 73\%$, $SD = 17\%$). However, neither of these comparisons was significant in Tukey HSD tests. Again, neither of the main effects was significant.

Reflecting the findings for the second and third hypotheses, in the overall $2 \times 2 \times 2$ ANOVA, there was a significant two-way interaction between self-motion and sound delivery, $F(1, 23) = 8.873$, $MSE = .009$, $p = .006$. No other effects were significant.

Signal Detection Analysis

Signal detection measures d' and C (See, Warm, Dember, & Howe, 1997) were calculated from the clicker data for each participant. For each measure, a 2×2 repeated-measures ANOVA was run with within-subjects factors of Self-Motion (walking vs. sitting) × Sound Delivery (earpiece vs. free field). Combined results are in Figure 4.

Signal detection analysis: d'. The main effect of self-motion was highly significant, $F(1, 23) = 25.645$, $MSE = .097$, $p < .001$, suggesting participants are less sensitive detectors of mismatches when walking than when sitting. There was also a nonsignificant trend for participants to be less sensitive detectors when using the earpiece than when using free-field sound, $F(1, 23) = 3.980$, $MSE = .066$, $p = .058$. A two-way interaction between self-motion and sound delivery, $F(1, 23) = 4.506$, $MSE = .141$, $p = .045$, indicated that walking reduced sensitivity compared with sitting much more when participants used the earpiece ($p = .001$) than when they used free-field speakers ($p = .475$).

Signal detection analysis: C. All means were greater than 0 and so showed conservative responding (relative reluctance to report a mismatch). There were no main effects of self-motion ($p = .438$) or for sound delivery ($p = .745$). However, there was a significant interaction of self-motion and sound delivery, $F(1, 23) = 5.651$, $MSE = .018$, $p = .026$, with responding relatively more conservative when participants were walking with the earpiece and when sitting with the free-field speakers, but no comparisons were significant in Tukey HSD tests.

Button-Press Results

Button-press accuracy. Because of the consistently high level of accuracy (range 94% to 100% per condition), a Friedman ANOVA was used to analyze the data. As predicted, there was no

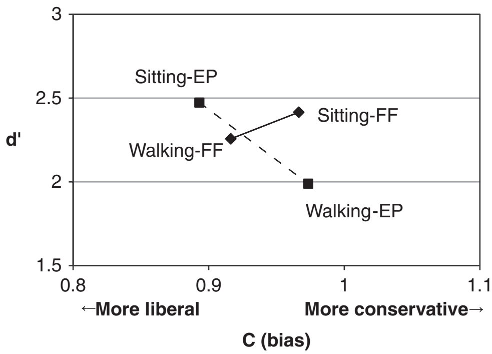

Figure 4. Signal detection results for mismatch identification with the clicker. The C measure of bias is shown in the *x*-axis. All values of *C* greater than 0 are conservative, so graph shows directions of relatively liberal ("yes") versus relatively conservative ("no") responding. The average *d′* results are shown in *y*-axis. EP = earpiece, FF = free field.

difference in how accurately participants pressed buttons across conditions, $\chi^2(31, 24) = 35.756$, $p = .429$, *ns*.

Button-press latency. Participants pressed buttons faster when they were sitting ($M = 1{,}791$ ms, $SD = 350$ ms) than when they were walking ($M = 4{,}328$ ms, $SD = 597$ ms), $F(1, 23) = 849.111$, $MSE = 1{,}455{,}618$, $p < .001$. Participants pressed the *A* and *B* buttons approximately 250 ms faster ($M = 2{,}938$ ms, $SD = 443$ ms) than the *C* and *D* buttons ($M = 3{,}182$ ms, $SD = 465$ ms), $F(3, 69) = 17.451$, $MSE = 244{,}248$, $p < .001$, all comparison *p*s < .01. Buttons *C* and *D* were pressed approximately 2,700 ms slower when participants were walking ($M = 4{,}527$ ms, $SD = 652$ ms) than when they were sitting ($M = 1{,}838$ ms, $SD = 378$ ms), $F(3, 69) = 14.148$, $MSE = 187{,}267$, $p < .001$, probably because the *A* and *B* buttons were located at the front of the room, whereas the *C* and *D* buttons were at the back, usually behind the participant.

Questionnaires

Significant results were found only for the fourth question, which probed how easy it was to time-share monitoring for mismatches with pushing the buttons. There was a significant effect of self-motion, $F(1, 23) = 14.748$, $MSE = 1.147$, $p < .001$, with participants rating time-sharing easier when they were sitting ($M = 4.104$, $SD = 1.343$) than when walking ($M = 3.510$, $SD = 1.307$), and of response method, $F(1, 23) = 5.146$, $MSE = 0.446$, $p = .033$, with participants rating timesharing easier when they kept a mental count ($M = 3.917$, $SD = 1.318$) than when they used the clicker ($M = 3.698$, $SD = 1.264$).

Sound delivery interacted with self-motion, $F(1, 23) = 5.447$, $MSE = 0.310$, $p = .029$, with the negative effects of walking rated greater for free field than for earpiece. Sound delivery also interacted with response method, $F(1, 23) = 7.290$, $MSE = 0.258$, $p = .013$, with the negative effects of walking rated as greater for the clicker than for the mental count. When walking, participants rated time-sharing most difficult when they were listening to free-field speaker sound and using the clicker. No other effects were significant.

The NASA-TLX did not show significant results that were specific to response method apart from a three-way interaction between self-motion, sound delivery, and TLX subscale that was largely uninterpretable.

DISCUSSION

Our results are unexpected in several ways. In this section, we review and interpret the findings. Then we outline possible explanations for the findings and suggest future studies. Finally, we point to practical implications.

Our first hypothesis was that keeping a count of mismatches with a clicker, rather than a mental count, would reduce the load on working memory and so improve mismatch task accuracy. Instead, the clicker worsened mismatch task accuracy. Moreover, participants reported that when using the clicker, it was harder to time-share monitoring for mismatches with pushing the buttons rather than easier.

This finding runs counter to the predictions of our extension to the Carlson and Cassenti (2004) model. First, the manual components of the clicker response may have interacted with manual and motor components of the button-press task. The demands of scheduling two manual tasks may have led participants to shift visual attention away from the HMD display more often than in other conditions and so occasionally to misidentify mismatches (Shin & Rosenbaum, 2002). Second, knowing that they could not "withdraw" a clicker response, participants may have adopted a more conservative response strategy overall, which is party supported by the generally conservative response bias with the clicker. However, neither suggestion explains why, specifically when walking with the clicker, participants count mismatches least accurately and with least sensitivity with the earpiece than with the free-field sound (see the third hypothesis).

Our second hypothesis was that we would replicate the findings of Thompson and Sanderson (2008) for conditions in which participants kept a mental count. Thompson and Sanderson found that with the earpiece, mismatch accuracy was unaffected by walking, whereas with free-field sound, accuracy was worsened by walking. The present results for the mental count do not replicate this finding. There is a trend for mismatch accuracy to be worse when participants are walking with the earpiece and worse when participants are sitting with free-field sound, but the interaction is not significant.

Our third hypothesis was that the clicker would remove the pattern of results found in Thompson and Sanderson (2008). Surprisingly, mismatch task accuracy and sensitivity (*d*) both showed a fully significant interaction in the opposite direction from that of Thompson and Sanderson when the clicker was used. Mismatch accuracy was worse when participants walked and used an earpiece and worse when they sat and used free-field sound. Response bias (*C*) indicated that worse performance was associated with more conservative responding and followed the pattern of results for mismatch accuracy. Nonetheless, participants' subjective reports indicated that they found time-sharing harder when they were walking and listening to free-field speakers.

These results are not caused by performance trade-offs with the time-shared button box task. Participants performed equally well at the button box task across all conditions. Participants were slower to push the buttons when they were walking than when they were sitting, simply because it takes time to walk to a button, but they were not slower to push the buttons under any combination of self-motion and sound delivery.

Interpretations and Future Directions

In what follows, we offer two conjectures as possible interpretations for the findings. Both indicate areas for future research.

The first conjecture is that theories of auditory representations in peripersonal space (Làdavas, Pavani, & Farnè, 2001) might help us construct an explanation for participants' worse accuracy at the mismatch task when they walk, hear bounce sounds with the earpiece, and use the clicker, as in the current study. An earpiece or headphone projects sound into the near, or peripersonal, space (Kallinen & Ravaja, 2007). The clicker may also influence spatial attention to the peripersonal space around the hand (Reed, Grubb, & Steele 2006). In addition, walking imposes greater attentional load than sitting (Lajoie et al., 1993), perhaps because of constant updating of the body's location in peripersonal space. With these three factors combined, walking with an earpiece and using the clicker may overload resources used to monitor peripersonal space. This would have led participants to miss some mismatches (as reflected in the more conservative response bias), leading to worse accuracy and lower sensitivity.

The second conjecture is that because the present experiment used a within-subjects design, in which participants experienced both mental count and clicker conditions, the results for the mental count may have been influenced by strategies that participants developed to manage the clicker. For example, in a mental count condition after a clicker condition, participants may have continued to tap their finger to help them keep count. Possible contamination by such carryover effects could be checked with a between-subjects manipulation of response method.

The previous two conjectures could be tested with a more sensitive measure of mismatch detection, such as a vocal response (Wickens, 2002). As noted in the Introduction, vocal responses may mask the auditory stimuli that are the focus of the study, but such masking would be approximately equivalent across combinations of sound delivery method and participant mobility. If the original Thompson and Sanderson (2008) results were replicated with a vocal response, then our conjectures about peripersonal space and carryover effects with the clicker would gain some credence. Theories of how mobile workers process multimodal input would then need to take response modality into account to provide accurate predictions. Such a finding would have important implications for the use of other modalities, such as haptic displays

(Ferris & Sarter, 2009; Ford et al., 2008; Ng, Man, Fels, Dumont, & Ansermino, 2005).

Further research is needed on the potential impact of sound delivery methods and worker mobility on the use of multimodal displays, including wearable computing. Moreover, it is important to move beyond the mismatch task used here to other tasks that capture how anesthesiologists, surgeons, pilots, soldiers, and others work with multimodal information. Exploring other modalities (e.g., haptic) not only would be useful for researchers and designers but also would provide important conceptual leverage in theory development.

Practical Implications

Thompson and Sanderson (2008) and Harrison et al. (2010 [this issue]) provide some evidence that workers using free-field sound report anomalies in multimodal object behavior less accurately when they are mobile and have high mental demands, but the present results suggest that their conclusion does not extend to all response modes.

At present, there is no theoretically coherent and empirically validated set of principles to help designers decide how to deliver auditory information to mobile workers. A review by Valjamae (2009) has made it clear that even in basic laboratory studies of multisensory interaction, barely any attention has been paid to the motion of participants. In our study, whether walking compromises performance on multisensory integration tasks and other tasks involving multimodal information appears to depend on how sound is delivered and on how responses are made. Moreover, whether the mobile worker benefits or suffers from the use of an earpiece versus free-field sound also seems to depend on concurrent mental and manual activities. Complicating the picture is the fact that many work contexts involve high mental and manual demands at different points in time. For example, at times, an anesthesiologist monitors a set of vital signs to determine a patient's state and, at other times, performs manual intubations or inserts central lines into a patient's venous or arterial system. One method of sound delivery may not suit all circumstances, and finding methods and means for changing sound delivery to suit the circumstances poses further design issues.

Multiple resource theory (Wickens, 2002) provides solutions to many multimodal design issues, but it does not readily distinguish the resource demands of different sources of auditory information. Until strong theory emerges about the effects of worker mobility and the differences between different methods of delivering information within a modality, rather than across modalities, designers are best advised to exercise caution and to rely on comprehensive formative and summative testing of designs involving such factors.

CONCLUSION

Overall, it is evident that when people are performing multiple multimodal tasks, relatively simple rearrangements of experimental or working conditions can lead to unpredictable and apparently contradictory results. Moreover, it is evident from this study and from both previous studies that conclusions drawn from investigating seated workers will not necessarily generalize to mobile workers (Harrison et al., 2010 [this issue]; Thompson & Sanderson, 2008). Complicating the picture is the potential influence of experimental design, in which results from one experiment will not necessarily generalize to another. A participant's strategy for performing in any one condition may reflect his or her reaction to other conditions. When asked to perform high-workload multisensory integration tasks, participants may be uniquely sensitive to the range of demands made across all conditions, leading to complex patterns of findings.

In their work on multimodal displays, Ferris and Sarter (2008) have commented that it is difficult to extrapolate findings from studies of multimodal processing from basic laboratory studies to the field. We venture that it may be difficult to extrapolate from any multimodal context, basic or applied, to any other multimodal context until broader theories are developed that cover all representative task elements and that are tested in a way that reduces undue sensitivity to specific experimental arrangements. Our study has reduced the possibility that simple, strong factors are at work. Discovering the more subtle factors that are at work is a challenge for advanced display studies addressing conditions likely to be encountered in the field.

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 ND2000 HMD that allowed us to complete the research. Finally, we acknowledge the help of members of the Cognitive Engineering Research Group 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). Oxford, England: Oxford University Press.

Carlson, R. A., & Cassenti, D. N. (2004). Intentional control of event counting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1235–1251.

Ferris, T. K., & Sarter, N. B. (2008). Cross-modal links among vision, audition, and touch in complex environments. *Human Factors*, *50*, 17–26.

Ferris, T. K., & Sarter, N. B. (2009). Supporting anesthetic monitoring through tactile display of physiological parameters. In *Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting* (pp. 1664–1668). Santa Monica, CA: Human Factors and Ergonomics Society.

Ford, S., Daniels, J., Lim, J., Loval, V., Dumont, G., Schwartz, S. K. W., & Ansermino, J. M. (2008). A novel vibrotactile display to improve the performance of anesthesiologists in a simulated critical incident. *Anesthesia and Analgesia*, *106*, 1182–1188.

Georgiou, P., & Kyriakakis, C. (1999). Modelling of head related transfer functions for immersive audio using a state-space approach. *Conference Record of the Asilomar Conference on Signals, Systems, and Computers*, *1*, 720–724.

Gray, R., Tan, T. Z., & Young, J. (2002). Do multimodal signals need to come from the same place? Cross-modal attentional links between proximal and distal surfaces. In *Proceedings of the 4th IEEE International Conference on Multimodal Interfaces* (pp. 438–442). Washington, DC: IEEE Computer Society.

Harrison, W. J., Thompson, M. B., & Sanderson, P. M. (2010). Multisensory integration with a head-mounted display: Background motion and sound motion. *Human Factors*, *52*, 78-91.

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 (Edsx), *Human Menta Workload*. (pp. 139-183). Amsterdand: North-Holland.

Ho, C., & Spence, C. (2008). *The multisensory driver: Implications for ergonomic car-interface design*. Hampshire, England: Ashgate.

Ho, C., & Spence, C. (2009). Using peripersonal warning signals to orient a driver's gaze. *Human Factors*, *51*, 539–556.

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*(1), 303–317.

Làdavas, E., Pavani, F., & Farnè, A. (2001). Auditory peripersonal space in humans: A case of auditory-tactile extinction. *Neurocase*, *7*(2), 97–103.

Lajoie, Y., Teasdale, N., Bard, C., & Fleury, M. (1993). Attentional demands for static and dynamic equilibrium. *Experimental Brain Research*, *97*(1), 139–144.

Laramee, R. S., & Ware, C. (2002). Rivalry and interference with a head-mounted display. *ACM Transactions on Computer-Human Interaction*, *9*(3), 238–251.

Liu, D., Jenkins, S., & Sanderson, P. M. (2009). Clinical implementation of a head-mounted display of patient vital signs. In *International Symposium on Wearable Computing (ISWC2009)* (pp. 47–54). Washington, DC: IEEE Computer Society.

Liu, D., Jenkins, S., Sanderson, P. M., Watson, M. O., Russell, W. J., Leane, T., & Kruys, A. (2009). Monitoring with head-mounted displays: Performance and safety in a full-scale simulator and part-task trainer. *Anesthesia and Analgesia*, *109*(4), 1135–1146.

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

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.

Ng, J. Y. C., Man, J. C. F., Fels, S., Dumont, G., & Ansermino, J. M. (2005). An evaluation of a vibro-tactile display prototype for physiological monitoring. *Anesthesia & Analgesia*, *101*, 1719–1724.

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.

Reed, C. L., Grubb, J. D., & Steele, C. (2006). Hands up: Attentional prioritization of space near the hand. *Journal of Experimental Psychology*, *32*(1), 166–177.

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*(6), 1787–1797.

See, J. E., Warm, J. S., Dember, W. N., & Howe, S. R. (1997). Vigilance and signal detection theory: An empirical evaluation of five measures of response bias. *Human Factors*, *39*, 14–29.

Shin, J. C., & Rosenbaum, D. A. (2002). Reaching while calculating: Scheduling of cognitive and perceptual-motor processes. *Journal of Experimental Psychology*, *131*(2), 206–219.

So, R. H. Y., Leung, N. M., Braasch, J., & Leung, K. L. (2006). A low cost, non-individualized surround sound system based upon head related transfer functions: An ergonomics study and prototype development. *Applied Ergonomics*, *37*(6), 695–707.

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

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

Watson, M. O., & Sanderson, P. M. (2004). Sonification helps eyes-free respiratory monitoring and task timesharing. *Human Factors*, *46*, 497–517.

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

Welford, A. T. (1980). Choice reaction time: Basic concepts. In A. T. Welford (Ed.), *Reaction times* (pp. 73–128). New York, NY: Academic Press.

Wenzel, E. M., Arruda, M., Kistler, D. J., & Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. *Journal of the Acoustical Society of America*, *94*, 111–123.

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

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

Morgan J. Tear is a researcher with National ICT Australia at the Queensland Research Laboratory in St Lucia, Queensland. He obtained his bachelor of psychological science in 2008 from The University of Queensland in St. Lucia, Queensland, Australia.

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 24, 2010