Multisensory integration with a head-mounted display and auditory display

Matthew B. Thompson and Penelope M. Sanderson
School of Psychology and School of Information Technology and Electrical Engineering
The University of Queensland

Human operators who use head-mounted displays (HMDs) in their work may benefit from auditory support. It is unclear whether auditory support is better delivered in free-field or via earpiece, and what the effect of walking is. To examine this problem, a novel multisensory integration task was created in which participants identified mismatches between sounds and visual information on an HMD. Participants listened to the sounds either via earpiece or free-field while they either sat or walked about the test room. When using an earpiece, participants performed the mismatch task equally well whether walking or sitting, but when using free-field sound, participants performed the task significantly worse when walking than when sitting. The worse performance for participants using free-field sound while walking may relate to spatial and motion inconsistencies between the sound and vision or because of misperceptions of the time at which the sounds occurred. The results underscore the need for representative design of experiments exploring multisensory integration and they suggest auditory conditions that might influence effective multisensory integration with HMDs.

INTRODUCTION

In the research reported here, we test 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, so making information continuously available to a mobile user (Patterson, Winterbottom, & Pierce, 2006). 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 (Brewster, 1994; Watson & Sanderson, 2004, 2007). Although there is much research on HMDs and on auditory displays alone, there is less research on how the two kinds of displays might work together, especially when people walk about when doing their work.

We present an initial study on the role of both self-motion and sound delivery methods on people's ability to perform multisensory integration of auditory information with visual information from an HMD.

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 the presence or absence of motion (walking, standing, sitting) might affect people's ability to perform multisensory integration with an HMD. In the very few studies comparing sound delivery methods, researchers have studied headphone vs. free-field delivery of news stories (Kallinen & Ravaja, 2007), the effect of headphone vs. free-field music on attention (Nelson & Nilsson, 1990), the effect of 2-D vs. 3-D headphone auditory support on people using an HMD to navigate in virtual space (Viaud-Delmon, Warusfel, Seguelas, Rio & Jouvent, 2006), and spatialization of sound to support visual tasks (Bolia, D'Angelo & McKinley 1999). No studies, however, have compared the effectiveness of different methods of sound delivery when a participant is moving.

Further, no studies have directly examined the effect of self-motion on HMD use. Laramee and Ware (2002) show that participants using a transparent HMD find it harder to perform visual tasks against a dynamic visual background, but the participants were seated. Studies have shown that self-motion can introduce attentional demands (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 over others.

Experimental Task

For our initial investigation we decided to test the effect of sound delivery and self-motion with a potentially very sensitive task and subsequently to move on to other tasks. We developed the so-called 'mismatch' task which requires multisensory integration. The task is based loosely on the Michotte (1963) launch task. In the 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 within the time window strengthens the perception of causality and extends the time window over which it occurs (Guski & Troje, 2003).

In our mismatch task, participants 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 mismatch.

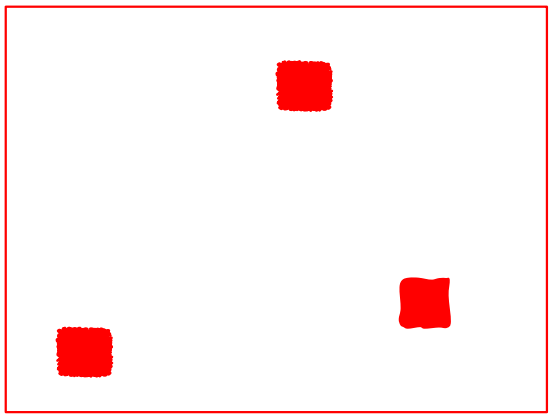

Figure 1. Mismatch task as displayed on the monocular, see-through HMD. Hard object is bottom right and has smooth contoured edges. Soft objects are at top and at left, and have "woolly" edges. Wall is considered a hard object.

Participants must integrate the visual and auditory information to detect whether the information matches or mismatches. Because visual and auditory information must be integrated to detect mismatches, factors that promote multisensory integration should improve performance.

Integration of multisensory information

According to Newell (2004), for multisensory object integration to occur information should be (1) task-relevant, (2) temporally congruent and (3) spatially congruent. The elements of our mismatch task are *task-relevant* because information from both modalities is required for a mismatch judgment, regardless of self-motion or sound delivery. The sound and vision of mismatch task collisions occur sufficiently close in time to seem *temporally congruent*, whether the sound is delivered via earpiece or via free-field (for tolerances see Burr & Alais, 2006; Lewald & Guski, 2004). In fact, temporal tolerances can be quite large, particularly if stimulus arrangements suggest a cause effect relationship between the vision and the sound.

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 others. Newell's (2004) spatial congruence criterion for multisensory integration therefore may not be met. We know that performance suffers when event sound and vision come from different spatial locations (Driver & Spence, 1994; Spence & Read, 2003). When a person using an HMD walks while listening to free-field sound, quite often there will be spatial incompatibility between the locations of object collisions in the visual field and associated collision sound in free-field. Similarly, there will often be motion incompatibility when the object appears to move in one direction but because of the person's motion, the sound appears to move in the opposite direction. Such incompatibilities may make integration more difficult. In addition, given what is known about different ventriloquism effects and the role of stimulus uncertainty (Ernst & Banks, 2002), the movement of the free-field collision sound may even make the sound harder to locate in time, and allow the wrong visual event to be associated with the wrong sound (Jaekl & Harris, 2007; Morein-Zamir, Soto-Faraco, & Kingstone, 2003).

Hypotheses

The first hypothesis is that when participants are walking, they will detect mismatches less accurately than when they are sitting, regardless of whether sound is delivered via free-field or earpiece. This is due to (1) visual interference from the background (Laramee & Ware, 2002) and (2) 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.

METHOD

Participants

The study was approved by The University of Queensland's School of Psychology ethics committee. Participants were 20 first-year psychology students aged between 17 and 39 (M = 22.3, SD = 5.6) who earned course credit for participation. All reported normal or corrected-to-normal vision and normal hearing.

Design

All participants experienced all eight conditions created by crossing Movement (Walking vs. Sitting), Sound Delivery (Free-field vs. Earpiece) and Mismatch Rate (Low vs. High). The order of presentation of conditions was counterbalanced to remove carryover or practice effects.

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. 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. The chin rest served both to limit motion of the HMD across the background and to limit changes in sound azimuth.

Sound delivery. In the Free-field condition, the sounds associated with object collisions came from a speaker in the corner of the test room. In the Earpiece condition, the sounds came from an earpiece placed in the participant's right ear.

Trials. In each experimental condition there were two separate trials of four minutes each, one with a relatively high rate of mismatches and the other with a relatively low rate. Low mismatch rates were 22-27 mismatches per scenario (7.4%-8.6% of total bounces) and high mismatch rates were 33-37 per scenario (10.3% to 11.3% of total bounces). Over the whole experiment, each experimental condition was observed for all scenarios and in all serial positions.

Figure 2. Quad display of a participant during the experiment. Participant is wearing the HMD and earpiece and is walking towards a pushbutton. Pushbuttons were located in each corner of the room. Letter "D" in frame 2 is a visual stimulus telling the participant which button-box to push.

Tasks

Multisensory object integration 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 being 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 heard 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)
- Soft object 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 (see Figure 2). Four button-boxes labeled A, B, C and D were placed in each corner of the room for the walking condition and in a similar configuration in front of the participant for the sitting condition. A large letter indicating the button-box to press was displayed on a computer screen at the front of the room. Participants' task was to press the button-box corresponding to the letter displayed on the computer screen. A notification sound alerting participants to the change came from a second loudspeaker and was a high-pitched, multiple-toned sound with long attack and decay that was distinctly different from the bounce sound.

Apparatus

The HMD was a Microvision Nomad™ ND2000 with a single optical see-through monocle (800 x 600). In the free-field condition, collision sounds were sent to a loudspeaker (Edirol™ MA-7A) using a wireless transmitter (Sony™ URX-P1/UTX-B1). The sound pressure level of the sounds, measured from the centre of the experiment room, was 70dB(A) max. In the earpiece condition, a single ear bud in the right ear with personal volume control was used (Sony™ MDR-E829V). The sound pressure level of the notification sound from the button-press task loudspeaker (Harmon/Kardon™ H/K695-04) was approximately 64dB(A) max.

Procedure

The volume of the free-field sound and the earpiece were equated as follows. The free-field sound was used as the standard and the participant adjusted the earpiece to match it. 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 & Rash, 1990).

Participants then learned to do the mismatch task and the button-press task together. First, the button-press task started and the objects appeared on the HMD, but the mismatch task sound was not played. Participants were asked to do the button-press task as efficiently as possible while making the mismatch task their primary task. They were asked to face each button-box when they pressed its button. Second, participants completed a full training scenario while sitting with an earpiece followed by a full training scenario while walking with free-field sound, using the mismatch task sounds in both cases. Then they proceeded to the experimental trials.

RESULTS

A 2x2x2 repeated measures ANOVA was conducted on the mismatch count 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 participants counted mismatches, $F(1,19) = 8.887$, $p = 0.008$, (see Figure 3). Participants counted mismatches more accurately when they were sitting than when walking. There was no main effect of Sound Delivery. There was a significant main effect of Mismatch Rate, $F(1,19) = 78.043$, $p < 0.001$. Participants counted mismatches more accurately when the Mismatch Rate was Low than when High (not shown in the figure).

As predicted, there was a significant two-way interaction between Movement and Sound Delivery, $F(1,19) = 6.449$, $p = 0.020$, (see Figure 3). However, the results of planned comparisons comparing Earpiece vs. 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 than with sound delivered through an earpiece, $p = 0.020$. When walking, however, participants counted mismatches equally well with sound delivered in free-field or earpiece.

A post-hoc Tukey HSD test 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 ($M = 74.7\%$, $SD = 19.6\%$), $p = 0.006$. No other post-hoc contrasts were significant.

As intended, button-press accuracy was extremely high in all conditions, ranging from 99.3% correct to 99.8% correct, so no inferential tests could be run. There was therefore no evidence of a tradeoff in accuracy between tasks across the walking and sitting conditions.

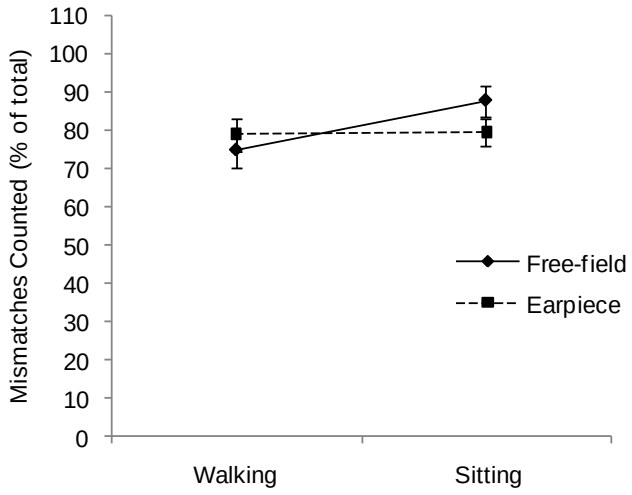

Figure 3. Mismatch Detection Accuracy (+ *SE*) for Movement (Walking vs. Sitting) and Sound Delivery (Free-field vs. Earpiece) conditions.

DISCUSSION

Our first hypothesis was that participants would detect mismatches less accurately when they were walking than when they were sitting, regardless of sound delivery method. This hypothesis was supported, but it is qualified by the interaction between movement and sound delivery.

Our second hypothesis was that when participants were sitting they would count mismatches equally well with earpiece and free-field sound, but when walking they would detect mismatches less accurately with free-field sound than with an earpiece. This hypothesis was not supported. Instead, the results are best characterized as follows. When sound came from an earpiece, participants counted mismatches equally well whether they were sitting or walking, which would not be predicted from the findings of Laramee and Ware (2002) and Sparrow et al. (2002). When sound came from free-field, however, participants counted mismatches less accurately when walking than when sitting.

The above result cannot be due to walking alone, because 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, the continuously changing spatial relationship between head-referenced HMD vision and world-referenced free-field sound may have made it more difficult for people to integrate sound and vision, and so to count mismatches accurately when walking. 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 (e.g. Jaekl & Harris, 2007; Morein-Zamir, et al., 2003).

This speculation should be pursued in future basic and applied studies, because it suggests one mechanism 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 reported here. 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 via a button-click so a signal detection analysis can be performed, but only if such a button-click response does not fundamentally alter the task.

Second, the effect of sound delivery method and self-motion on tasks that can be performed with vision or sound alone ('redundant' tasks) rather than requiring both ('integration' tasks, as here) is still unknown.

Third, participants performed a task involving a dynamic display on the HMD. It is unknown how a task involving a static visual display on the HMD might fare (such as that used by Laramee & Ware, 2002) with earpiece vs. free-field delivery of sound.

Fifth, overall performance with the earpiece was not as good as expected. The apparent location of the sound (towards the right ear) may not have been the best match for the apparent central location of the visual display and so may not have provided best conditions for multisensory integration. Placing the perceived location of the sound in the centre of the head, possibly by using headphones, may increase spatial congruence and improve performance.

Conclusions

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. Most prior research on motion and auditory stimuli has examined the perception of moving auditory stimuli, rather than the perception of auditory stimuli by a moving listener. No research has investigated factors influencing people's ability to perform multisensory integration across the visual and auditory modalities while people walk about, yet it is a common human task. New display technologies underscore the need for research that extends current theories of multisensory information processing and ecological perception.

ACKNOWLEDGEMENTS

This research was supported by Australian Research Council grant DP0559504 to Penelope Sanderson, Marcus Watson, and W. John Russell. We thank Dr Stas Krupenia and David Liu for their pioneering work in our laboratory on HMDs and for their help with this project.

REFERENCES

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*(3), 257-262.

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: US Army Aeromedical Research Laboratory, Fort Rucker, AL.

Bolia, R. S., D'Angelo, W. R., & McKinley, R. L. (1999). Aurally aided visual search in three-dimensional space. *Human Factors, 41*(4), 664-669.

Brewster, S. A. (1994). *Providing a structured method for integrating non-speech audio Into human-computer interfaces.* Ph.D dissertation. University of York, York, UK.

Burr, D., & Alais, D. (2006). Combining visual and auditory information. In S. Martinez-Conde, S. L. Macknik, L. M. Martinez, J. M. Alonso & P. U. Tse (Eds.), *Progress in Brain Research* (Vol. Volume 155, Part 2, pp. 243-258): Elsevier.

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

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*(6870), 429-433.

Guski, R., & Troje, N. F. (2003). Audiovisual phenomenal causality. *Perception & Psychophysics, 65*(5), 789-800.

Jaekl, P. M., & Harris, L. R. (2007). Auditory–visual temporal integration measured by shifts in perceived temporal location. *Neuroscience Letters, 417*(3), 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.

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

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*(2), 119-122.

Michotte, A. (1963). *The perception of causality. Translated by T. Miles and E. Miles*: Basic Books.

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Cognitive Brain Research, 17*(1), 154-163.

Nelson, T. M., & Nilsson, T. H. (1990). Comparing headphone and speaker effects on simulated driving. *Accident Analysis and Prevention*, 22(6), 523-529.

Newell, F. N. (2004). Cross-modal object recognition. In G. A. Calvert, C. Spence & B. E. Stein (Eds.), *The Handbook of Multisensory Processes*: Bradford Books.

Patterson, Winterbottom, M. D., & Pierce, B. J. (2006). Perceptual issues in the use of head-mounted visual displays. *Human Factors, 48*(3), 555-573.

Sparrow, W. A., Bradshaw, E. J., Lamoureux, E., & Tirosh, O. (2002). Ageing effects on the attention demands of walking. *Human Movement Science, 21*(5), 961-972.

Spence, C., & Read, L. (2003). Speech shadowing while driving: on the difficulty of splitting attention between eye and ear. *Psychological Science*, 14(3), 251.

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.

Watson, M. O., & Sanderson, P. (2004). Sonification Supports Eyes-Free Respiratory Monitoring and Task Time-Sharing. *Human Factors, 46*(3), 497-518.

Watson, M. O., & Sanderson, P. (2007). Designing for attention with sound: Challenges and extensions to ecological interface design. *Human Factors, 49*(2), 331-346.