Crossmodal integration with a head-mounted display and auditory display options:

Is there cause for concern?

Crossmodal integration with a head-mounted display and auditory display options: Is there cause for concern?

Matthew Thompson

BSc (Psychology), BInfTech

School of Psychology

University of Queensland

This thesis is submitted to the School of Psychology in partial fulfilment of the requirements of a Bachelor of Science (Honours) degree.

Date: 16th October 2007

Word Count: 14 500

Statement of Originality

I, Matthew Thompson, certify that this thesis has not been submitted previously for the requirements of any university degree. The thesis comprises my own work, except where reference is made in the text itself.

Matthew Thompson

16th October 2007

Acknowledgements

Penny, I am indebted to you for the wisdom, support, guidance and motivation you provided this year. I have learned a great deal.

I would like to issue a formal apology to my friends and family for my unusual, erratic and generally unpleasant behaviour during the course of this thesis. Your support is greatly appreciated.

Thankyou to the members of the Cognitive Engineering Research Group for your friendship and helpful feedback: David Liu, Tania Xiao, Marcus Watson, Vivian Shek, Xilin Li, Tobias Grundgeiger, Sean Lowe, Phil Cole, Stas Krupenia, David Crone and Itsik Nadler.

I owe a special thankyou to Stas Krupenia for being available for discussion at the drop of a hat. You taught me heaps Stas and I have enjoyed many great times with you.

Thankyou to the following people for giving up their time to help conceptualise this thesis: Professor David Crewther, Dr Derek Arnold, Professor Jason Mattingly, Dr David Alais, Dr Steve Barrass, Dr Dan Raemer, Professor Cathy Burns, Professor Denis Burnham, Professor Heinrich Bülthoff and Dr Simon Jenkins.

Finally, Jess, I hope one day to repay you for the unwavering support you have given me.

I love you.

Abstract

Head-Mounted Displays (HMDs) are increasingly used to support mobile work (Laramee & Ware, 2002). Human operators sometimes require additional auditory support when using an HMD, which raises the question of whether sound is better delivered publicly in free-field or privately via earpiece. A novel experimental procedure was created in which participants had to identify mismatches between auditory information and visual information on an HMD. Different conditions of sound delivery and physical movement were manipulated within-subjects. Participants heard the sound either via earpiece or free-field while they either sat or moved about the test room. Predictions were based on the idea that inconsistent spatial mapping of vision and sound would compromise mismatch detection. First, I predicted a main effect of movement such that participants' mismatch detection would be worse when they moved than when they sat. Second, I predicted an interaction between movement and sound delivery. When participants are seated there will be no difference in mismatch detection between the two methods of sound delivery. When participants are walking, however, mismatch detection will be better with an earpiece than with free-field delivery. Results supported the first prediction. For the second prediction, the significant interaction found took a different form than predicted. With the earpiece, participants performed equally well whether sitting or walking, but with free-field sound, participants performed better when sitting than when walking. Results have implications for understanding necessary auditory conditions for effective crossmodal integration and may indicate a cause for concern for people who use HMDs and auditory displays in safety-critical environments.

Publications

Thompson, M., Lowe, S., & Sanderson, P. (2007). Role of motion and sound in use of head-mounted displays (Abstract). Proceedings of the 8th International Multisensory Research Forum (IMRF2007). Sydney, 5-7 July, 2007.

CONTENTS

List of Figures

List of Tables

Crossmodal integration with a head-mounted display and auditory display options: Is there cause for concern?

The experimental design of the current thesis is motivated by the arrival of advanced visual and auditory displays in safety-critical environments and particularly to critical care. The specific case I will examine is motivated by a practical problem: how well people can integrate information from Head-Mounted Displays and auditory displays while they are moving around a room and listening to the auditory display either from an earpiece or from speakers in free-field? The thesis investigates crossmodal integration under more complex and naturalistic conditions than usually reported.

I will first describe the two technologies that will be the focus of the thesis, Head-Mounted Displays and auditory displays, and give some background on their use. The applied literature will then be examined for guidance on the practical problem outlined above. I will then turn to basic literature that investigates crossmodal integration to provide a basis from which to derive experimental hypotheses.

Motivation for the Research

Advanced display technologies are currently being researched and developed with the aim of supporting the work of human operators in safety critical environments (Grabowski & Sanborn, 2003; Sarter, 2006). Transparent Head-Mounted Displays (HMDs) are an advanced display currently receiving research attention. An HMD is a portable monitor worn on the operator's head that projects information onto a transparent screen (Melzer & Moffitt, 1997, see Figure 1). Information on an HMD is presented so it can be viewed simultaneously with other information in the environment (Yeh, Wickens, & Seagull, 1999). The information on the HMD is always available to the wearer, wherever they are looking in their environment.

Figure 1. Transparent Monocular Head-Mounted Display

Domains in which HMDs are currently being applied include aviation, industry and healthcare (Laramee & Ware, 2002; Ormerod, 2002; Perrott, Cisneros, McKinley, & D'Angelo, 1996). Researchers have identified advantages associated with the use of HMDs in enabling simultaneous viewing of multiple sources of information (Block, Yablock, & McDonald, 1995; Krupenia, 2007; Via et al., 2003; Yeh, Merlo, Wickens, & Brandenburg, 2003).

Despite the apparent advantages of HMDs, there are issues associated with their use in many domains (Keller, 1998; Patterson, Winterbottom, & Pierce, 2006). Specifically, the use of an HMD may make people more likely to miss unexpected events in the environment (Krupenia, 2007). Some researchers have suggested that transparent monocular HMDs are not suited for use in dynamic environments (Laramee & Ware, 2002). Laramee and Ware investigated two kinds of visual background onto which information from a transparent monocular HMD was projected. A dynamic background was created by a movie on a large screen TV and a static background was created by a stationary bookshelf. While seated, participants made a mouse-click response to the correct cell in a table. Participants were slower to respond when performing the task with the dynamic visual background than with the static visual background. Results suggest that the slower responding is due to visual interference that exists when information on an HMD is projected against a moving background. From these results Laramee and Ware posit that transparent monocular HMDs may not be suited for use in dynamic environments.

Auditory displays are another kind of advanced display receiving research attention in safety-critical domains (Kramer, 1994). Information from an auditory display is conveyed through sound rather than vision (Kramer, 1993) which is useful in work contexts in which vision is unavailable or overloadedFor example, a fighter pilot may rely on 3-D auditory cues about the location of enemy aircraft (Wickens, Goh, Helleberg, Horrey, & Talleur, 2003) or a stock market trader listens to a complex acoustic signal that tells him about the state of different financial commodities in the market while he performs complex visual tasks (Nesbitt & Barrass, 2002)). In medicine, an excellent examples is pulse oximetry where the rate of beeps conveys a patients' heart rate and the pitch of the beeps conveys their oxygen saturation level to caregivers who are often interrupted or distracted (Craven & McIndoe, 1999).The advantage of an auditory display is that information is always available, regardless of a listener's orientation in space. In particular, auditory displays have the potential to recede into peripheral awareness when all is well, but to capture attention when needed (Woods, 1995). When available to all members of a team, auditory displays can aid team communication and coordination (Patterson, Watts-Perotti, & Woods, 1999). Possible disadvantages of auditory displays in safety-critical environments is that they can be noisy, intrusive and may capture attention inappropriately if badly designed (Edworthy & Hellier, 2005; Seagull & Sanderson) In addition, there is evidence that different personalities tolerate ambient sound differently (Furnham & Strbac, 2002). These and further issues will be discussed in more detail shortly.

Researchers have started to examine the use of different kinds of advanced displays in the medical domain (Ormerod, 2002, 2003; Via et al., 2003). More recently, researchers have investigated the use of both HMDs and continuous auditory displays in support of anaesthetists’ patient monitoring tasks (Sanderson, 2006; Sanderson, Watson, & Russell, 2005) and have found that performance is best when both are used together (Sanderson et al., 2007). Auditory displays can provide continuous or discrete information to anaesthetists about patients’ vital signs (Seagull, Xiao, Mackenzie, & Wickens, 2000). It has been suggested that auditory displays should convey time-critical information and visual displays should convey less urgent information (Morris & Montano, 1996). However, researchers are only just starting to explore the potential for auditory displays, rather than auditory alarms or speech-based notifications, and to consider how best to integrate them with visual displays.

Practical Problems in Real-World Contexts

A set of problems has emerged from researchers’ exploration of HMDs and auditory displays in critical care environments that concerns the best way to provide information to carers using advanced technology (Sanderson, 2006). These problems may be equally relevant for other work domains. In critical care environments such as the operating theatre, medical staff can be physically oriented in any of several ways when doing their work. For example, a doctor can be sitting, standing or walking around. HMDs and auditory displays have the potential to support the mobile nature of doctors’ work. Moreover, especially with the advent of continuous auditory displays, the doctor’s attention can be guided by sound as well as by vision (Watson & Sanderson, 2004, 2007). Because information presented to people is becoming increasingly multimodal (Cohen & McGee, 2004; Sanderson, 2006; Sarter, 2006), it is important that sound and vision work effectively together when people are supported by advanced displays.

In the medical domain, some doctors have expressed concerns about introducing more sound into the already-noisy operating theatre (Hodge & Thompson, 1990; Kam, Kam, & Thompson, 1994). Accordingly, it has been suggested that sound should be personalised by being presented through an earpiece (Sanderson et al., 2007). However, team coordination in the operating theatre may be better achieved when sound is available to everyone - anaesthetists, surgeons, and nurses (Sanderson, 2006). Therefore, sometimes it may be beneficial to have a private auditory display and sometimes a public auditory display. If auditory information is delivered to practitioners privately via an earpiece, however, it is important that other sounds from the environment are also audible to them. Based on these considerations there are two ways to deliver sound using current technology: (1) publicly in the free-field via loudspeaker or (2) privately via personal earpiece.

To date (e.g. Sanderson, 2006), discussions about the advantages and disadvantages of public (free-field) vs. private (earpiece) form of sound delivery have been confined to their possible effects on teamwork and communication. Exactly how the form of sound delivery might affect cognitive and perceptual processes has not been addressed, yet there are areas of potential concern. First, the mobile nature of critical care may make performance of tasks that involve an HMD worse due to visual interference from background motion (Laramee & Ware, 2002). Second, when sound is delivered through loudspeaker to people who are working with HMDs in a mobile environment, there is potential for spatial mismatch between the location of the sound and the vision. These factors may affect people's ability to perform tasks when they are receiving information from more than one sensory modality. Specifically, I pose the question: Is there cause for concern for people using HMDs and auditory displays together under some conditions and might their performance even suffer?

To answer this question, evidence from applied studies will first be examined to determine how the method of sound delivery might affect people's performance while they are mobile. Evidence from studies of crossmodal perception and attention will then be examined in relation to the kind of task that will be explored in this study. It will be shown that neither the applied nor basic literature provides an answer to whether the method of sound delivery is critical when people are performing multimodal tasks in dynamic environments. The literature will, however, provide a theoretical basis from which to generate experimental hypotheses.

Evidence from Applied Studies

It would seem that the human factors and human movement literature should provide an answer to the question of how the method of sound delivery (earpiece, headphone, free-field, etc.) and presence or absence of motion (walking, standing, sitting, etc.) might affect people's ability to use auditory displays and HMDs. Accordingly, the applied literature will be reviewed and evaluated.

Studies relevant to the above questions can be classified into three broad categories, (1) studies in which researchers examine performance with different modes of sound delivery, (2) studies in which researchers examine how spatialisation of sound might support performance in visual tasks, and (3) studies in which researchers examine how human movement affects performance of perceptual and cognitive tasks. It will become clear, however, that different methods for delivering sound have not been investigated in a way that provides an answer to the above research question, or even in a way that provides a strong basis on which to make predictions. Research on spatialisation of sound is not informative about different methods of sound delivery for tasks that require integrating audio and vision, or about how sound delivery might be affected by movement. Finally, research on human movement does not inform us about people's ability to assimilate visual and/or auditory information.

Methods of Sound Delivery

In one of very few studies to directly compare two different methods of delivering sound, Kallinen and Ravaja (2007) investigated people's psychophysiological and subjective emotional responses to either headphones or loudspeakers when listening to news from a computer. The authors comment on the lack of research into the role of personal space in the study of human-computer interaction and take a first step to remedy this. Whether sound was presented through headphones or loudspeakers made no difference to participants' ability to comprehend news reports. However, participants made more positive emotional responses to headphones and preferred headphone to speaker listening. These results suggest that people may feel more strongly about how sound is delivered than is apparent in their performance. Because people only had to attend to information in the auditory modality, however, it is difficult to generalise these results to a task that requires both audio and vision.

The method of music delivery on driving performance was investigated in a driving simulator study by Nelson and Nilsson (1990) in which participants heard irrelevant music either through headphones or through a speaker. It was found that participants' gear shifts, their most complicated task, was delayed when they heard the music via headphones rather than via speakers. The authors suggest that headphones restrict auditory attention more than speakers do. It is difficult to generalise these findings to tasks in which sound and vision are both related to the task, but the fact that headphones apparently captured attention to irrelevant sound more strongly than speakers did suggests that headphones may also be more effective when both sound and vision are relevant to the task.

In a study investigating navigation with an HMD in a virtual world, Viaud-Delmon, Warusfel, Seguelas, Rio and Jouvent (2006) compared participants' performance with either visual information alone or auditory and visual information together. No difference in navigation performance between the two conditions was found. Visual information was presented on an HMD and sound was presented through headphones, but no other sound delivery method than headphones was used. The relative advantage or disadvantage of having sound delivered through headphones, therefore, cannot be directly examined. In addition, participants' moved through the virtual world only while standing, so the effect of movement cannot be determined.

This brief review has made it clear that different methods for delivering sound have not been directly investigated. Further, no studies have simultaneously examined the effect of participant movement with different methods of sound delivery on performance. Many studies, however, have examined the spatialisation of sound delivery, as reviewed below.

Vision and Spatialised sound

Most of the applied literature on the relationship between vision and sound is concerned with spatialisation of sound to support visual tasks. Paradigms usually involve a spatial audio cue to a visual event to be detected anywhere in 360 degrees. The research is particularly active in the area of aviation where the goal is to support pilots' ability to detect targets in the outside world. Studies often compare the effects of spatial audio presentation via loudspeakers or headphones (Bolia, D'Angelo, & McKinley, 1999; Perrott et al., 1996). One research goal is to use headphones to provide three-dimensional audio cues to visual target locations. This literature will be surveyed to determine whether it is informative about different methods of sound delivery for integrated tasks and how the effectiveness of different methods of sound delivery may be affected by movement.

In a flight simulator study, Begault (1993) presented participants with a sound cue either through one earpiece without spatial information or through two earpieces with additional 3D spatial information. Pilots' acquired targets significantly faster with two-earpiece 3D presentation. Although this study highlights possible benefits of adding redundant spatial information to an auditory display for a task that requires localisation, it does not tell us about potential differences between alternate methods of sound delivery for a task that requires integration of sound and vision.

In two studies similar to the one above, Perrott et al. (1996) and Bolia, D'Angelo and McKinley (1999) investigated the effectiveness of spatial audio displays on target acquisition performance. The authors compared three different sound delivery methods: no audio, free-field spatial audio with speakers, and simulated free-field audio with headphones. Both studies found that participants' performance was worse when sound was delivered through headphones than when sound was delivered through speakers. Because the same information was presented in both sound delivery conditions, these results seem to suggest that headphone and speaker listening are different in some way. The authors caution, however, that this difference is most probably due to technical limitations that meant that free-field cues were only imperfectly replicated in the headphone condition. The lack of accurate free-field cues in the headphone condition may have generated uncertainty for participants in localising targets and would explain their poorer performance. These studies are not completely helpful, therefore, in understanding differences between headphone and speaker sound delivery.

Effects of Walking vs.Sitting on Task Performance

Walking would seem to be a very low-workload task that would not absorb attentional resources and therefore would not affect people's ability to process visual and auditory information. Surprisingly, however, several human movement studies have used dual task methodologies to show that walking can absorb attentional resources and can affect people's ability to perform timeshared tasks (Abernethy, 1988; Sparrow, Bradshaw, Lamoureux, & Tirosh, 2002).

For example, Sparrow et al. (2002) showed that participants' response times to both auditory and visual secondary task probes slowed when the participants were walking compared to when they were standing still. The visual input required to perform the visual secondary task may have competed structurally with the visual input required to walk, but the auditory input required for the auditory task involved no such structural competition.

Therefore the auditory input must have competed for attentional resources with walking. Sparrow et al. (2003) concluded that that there was an attentional cost associated with walking and the increased attention demands of walking would reduce the resources available for other (secondary) tasks. Additionally, Lajoie, Teasdale, Bard and Fleury (1993) measured participants' reaction time to auditory probes occurring at unexpected intervals while the participants sat, stood, or walked. Results indicated that standing and walking involved more attentional demand than sitting.

The above studies indicate that walking absorbs more attentional resources than sitting, when measured by simple reaction time tasks. The studies do not indicate the degree to which walking vs. sitting will interfere with a more complex judgment task involving both auditory and visual information, nor do they provide a basis on which to predict whether walking would selectively affect one method of sound delivery (e.g., earpiece or free-field) over another.

Summary

In summary, applied studies have not directly examined the effects of sound delivery on such performance, and none has examined the effects of movement on such performance. Very few studies have directly compared different methods of sound delivery and the results of those that have are unclear. Studies on spatialisation of sound are not helpful in understanding how earpiece sound delivery might be different from free-field sound delivery. Finally, the effect of walking on simple reaction time tasks is known, but the effect of walking on more complex audio-visual tasks and on different methods of sound delivery has not been directly investigated. Much of the applied research reviewed deals only with visual tasks for which sound provides additional redundant cues, rather than tasks that require a judgment about the integration of vision and sound.

Overall, the applied literature does not provide an answer to the research question, nor does it provide a strong basis with which to make predictions. I must turn, therefore, to the more fundamental literature to find a basis on which to make predictions about the effect of methods of sound delivery and of movement on people's ability to perform tasks with HMDs and auditory displays. However, there are many different kinds of information that can be presented on an HMD and there are many different kinds of auditory displays. In addition, there is a vast array of tasks that people may be required to perform with HMDs and auditory displays. The search space created by the combination of these factors, in an attempt to cover all situations, is unreasonably large. The present thesis represents just the first step in an investigation of these factors--the first step of what might become a large and systematic program of research. Therefore it is necessary to limit the scope of the inquiry to a subset of visual and auditory information, and a subset of possible tasks. Accordingly, the next section describes the specific auditory and visual displays and the specific task that have been chosen as a starting point for this program of research. This makes it possible to identify which parts of the vast body of theoretical literature are potentially relevant and can form the basis for specific hypotheses.

The Experimental Task Chosen

An experimental task has been designed for the current thesis that captures some high-level properties of the tasks that people perform with auditory and visual displays in dynamic work environments. The task does not necessarily reflect any particular task that people currently perform. Instead it reflects an arrangement of visual and auditory stimuli that might be particularly demanding, and so create cause for concern. If I can establish that there are arrangements for using HMDs and auditory displays that compromise performance, and others that do not, then I can investigate why and make recommendations as to how to avoid such arrangements.

On the basis of substantial pilot research, I created an experimental task in which people must integrate information from both vision and sound (an “integration” task), rather than being able to perform the task on the basis of vision or sound alone (a “redundant” task). The task is based loosely on the Michotte launch task (Michotte, 1963) in which one object “hits” another and makes it move. People watching the Michotte launch task perceive a causal relationship between the first and second object if the second moves within a certain time window. If an auditory cue occurs at any point during the visual delay, however, it greatly strengthens the perception of causality and increases the time window over which the phenomenon occurs (Guski & Troje, 2003). I developed a continuous, two-dimensional version of the Michotte launch task and added auditory information about object behaviour. Participants monitor objects moving on an HMD as the objects collide and bounce off each other according to Newtonian laws of motion. Some of the objects are hard-looking and some are soft-looking (see *Figure 2*). When they collide, objects usually make a sound that is consistent with whether they look hard or soft. Participants must integrate the visual and auditory information about the objects to detect whether the visual and auditory information match or mismatch. This task is referred to as the ‘Mismatch task’ and it satisfies De Gelder and Bertelson’s (2003) call for greater naturalism and representativeness in experimental paradigms used to study multisensory integration. The issue under investigation is whether an integrated crossmodal task of this kind is performed better with an earpiece or with free-field sound delivery when participants, who are monitoring the objects on an HMD, are free to move around the room or are seated.

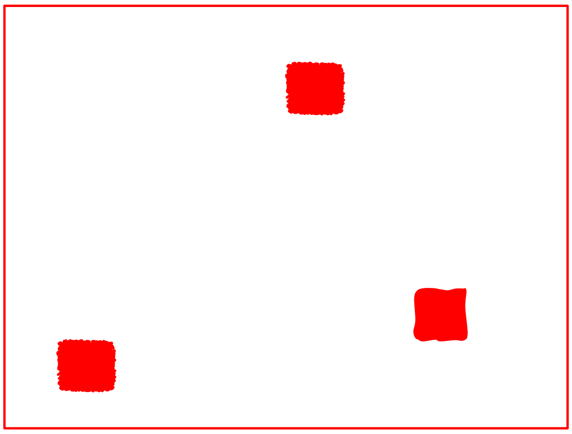


Figure 2. Moving Shapes on the HMD for the Mismatch task

Because both visual and auditory information is needed to detect mismatches in object behaviour, factors that promote multimodal integration should improve performance of the Mismatch task (Calvert, Spence, & Stein, 2004). For example, factors that help participants perceive the auditory and visual information as coming from a single object may improve mismatch detection performance. The literature on crossmodal integration, therefore, help determine whether earpiece or free-field sound delivery will be better when participants are walking around with an HMD. I focus on tasks involving crossmodal integration rather than on tasks that can be performed with vision or sound alone. In the latter tasks, sound and vision do not necessarily need to be integrated for the task to be performed, but performance may be enhanced by the use of both sound and vision. Such “redundant” crossmodal tasks, may be studied later in a continuation of the present research. To reiterate, the focus of the current thesis is on tasks that require people to integrate information from vision and sound sources in order to make judgments.

Crossmodal Perception and Attention

Given the absence of directly relevant applied studies, the first principles of human crossmodal perception and attention might provide a theoretical basis from which to generate experimental hypotheses. I review this literature for any evidence that multisensory integration may work differently in the mismatch task, described above, when sound is delivered via earpiece vs. free-field, or when people are walking rather than sitting. In what follows, I review research on multisensory integration, relating it to the questions underlying the present research. Finally, conclusions will be drawn from which specific experimental hypotheses are generated.

Integration of Crossmodal Information

There is a large body of research examining how humans integrate information about objects from different sensory modalities (e.g. Calvert et al., 2004; Spence & Driver, 2004; Stein & Meredith, 1993). When an object's properties are not completely specified from any single modality, information from other modalities is integrated to provide a complete representation of that object. In order to guide judgments about objects, information from sound and vision must converge to form a coherent percept (Newell, 2004).

Spence and Driver (2004) suggest that in many situations, people may benefit from combining information from different senses. For example, it is easier to listen to a person speaking in a noisy environment if their lip movements are visible (Sumby & Pollack, 1954). In this case, attending to the redundant visual stimulus of the lips promotes understanding of the auditory stimulus. Similarly, Giard and Peronnet (1999) found that when participants were presented with crossmodal information about an object, they were faster and more accurate at identifying the object than when presented with unimodal information. It appears that when two modalities provide independent pieces of information about an object, the information is combined in a statistically optimal fashion to facilitate recognition of its properties (Ernst & Banks, 2002). Recent evidence suggests that multisensory integration becomes much less effective if participants concurrently perform an unrelated visual or auditory task, indicating that multisensory integration is not automatic but instead demands attentional resources (Alsius, Navarra, Campbell, & Soto-Faraco, 2005). Overall, more efficient recognition of objects occurs when information about objects can be integrated across modalities (Newell, 2004).

For object integration to occur, Newell (2004) proposes that information should be (1) task-relevant, (2) temporally congruent and (3) spatially congruent. The following section will consider whether the Mismatch task meets the three preconditions for object integration. As will be seen, the elements of the Mismatch task are relevant and temporally congruent with each other, but under some of the conditions are not spatially congruent. This fact is important for the experimental predictions and for the findings of the thesis.

Task-relevance. The first requirement for crossmodal integration of vision and sound is that the information people receive from both modalities is relevant to their task (Newell, 2004). The Mismatch task has been constructed so that information from both the auditory and visual modalities must be used in order to make a judgement. The Mismatch task thus fulfils the first requirement of multimodal integration as the information in both modalities is relevant to the task. The task-relevance of the auditory and visual information in the Mismatch task will remain the same regardless of manipulations in sound delivery method or participant movement.

Temporal congruence. Temporal congruence between sound and vision is the second requirement for successful integration of crossmodal information (Newell, 2004). A time delay between an event occurring in one modality and an event occurring in another modality can reduce the likelihood of integration. Many laboratory studies have manipulated the temporal asynchrony between vision and sound and found that large stimulus-onset asynchronies (SOAs) make integration less probable (Lewald & Guski, 2003; Slutsky & Recanzone, 2001). Integration will still occur, however, for SOAs of less than 100ms. It has been demonstrated that auditory and visual information about the same object will be perceived as simultaneous even given time delays for arrival of sounds associated with distances of up 20 metres (Lewald & Guski, 2004). It may be that humans compensate for the physical delay introduced by the relatively slow speed of sound (Burr & Alais, 2006). As Burr and Alais argue, crossmodal integration is an active interpretive process capable of taking environmental cues into account. It is therefore unlikely, that any temporal asynchrony between sound and vision, introduced by either the manipulation of sound delivery method or movement, will affect the integration of sound and vision to create a perceptual object in the Mismatch task.

Spatial congruence. The third and final requirement for integration of crossmodal information is that sound and vision must be in the same spatial location; that is, they must be spatially congruent (Newell, 2004). Studies have demonstrated that presenting sound and vision in different spatial locations is detrimental to performance (e.g. Driver & Spence, 1994) and that stimulus motion can bias perception (Lewald, Ehrenstein, & Guski, 2001).

Driver and Spence (1994) found that, for a task that required crossmodal integration of information from vision and sound, people's performance worsened when the vision and the sound were in different spatial locations. Participants heard two verbal messages through loudspeakers from two spatially separate locations. Participants shadowed the target message while ignoring the distractor message. Screens of visual information showed lip movements that were either consistent with the shadowed message or unrelated (and meaningless). In the critical manipulation, the relevant lip movements were presented either near the target message or near the distractor message. Participants shadowed messages better when the relevant lip movements were near the target message. This suggests that when people perform a task that requires integration of vision and sound to make a judgement, their performance is worse when the vision and the sound are presented at different spatial locations.

The above cost to performance persists even when the visual and auditory modalities support separate tasks. Spence and Read (2003) extended Driver and Spence's (1994) findings in driving simulator. Participants performed an auditory speech shadowing task and a demanding driving task. The message to shadow was presented to participants through a loudspeaker placed either directly in front of them, or ninety degrees to their side. While driving and attending to the scene ahead, participants found it harder to shadow speech presented from their side than from directly in front of them. People may do worse because it is difficult to direct visual and auditory attention to different spatial locations (Driver & Spence, 1994).

Studies of the famous ventriloquist illusion also offer conditions under which vision and sound may or may not appear to come from the same object (Howard & Templeton, 1966). The ventriloquist illusion is most conventionally viewed as the tendency for people to mislocate sounds toward their apparent visual source (Alais & Burr, 2004; Bertelson, 1998). In the research paradigm, the spatial disparity between the auditory and visual source is manipulated. The more spatially separated the auditory and visual stimuli are, the less likely it is that the auditory stimulus will be captured by the visual stimulus (Jack & Thurlow, 1973; Slutsky & Recanzone, 2001; Thurlow & Jack, 1973), and the most reliable stimulus will capture the less reliable one (Alais & Burr, 2004).

Some tasks in the studies reviewed previously involve the integration of auditory and visual stimuli whereas others do not. In general, there is a performance cost when sound and vision are in separate locations (Driver & Spence, 1994; Howard & Templeton, 1966). In some situations this may be because it is difficult to attend to separate spatial locations across two modalities. In other situations, spatial disparity of information from vision and sound may not allow people to integrate the information to form a single perceptual object. The benefits of using crossmodal information are realised when sound and vision are presented at the same spatial location relative to each other.

Further studies indicate that motion can affect the spatial and temporal limits within which vision and sound are integrated to give the impression of coming from an object at a fixed location. For example, Lewald et al. (2001) present evidence that there is a spatio-temporal window for multimodal object integration that lasts about 100 msec in time and extends about 3 degrees in space. If visual and auditory stimuli fall outside these limits with respect to each other, the participant's sense of where they intersect in space will be biased. I conjecture that when bounce sounds move in space, as with walking, the subjective location at which visual and auditory properties of the bounce appear to coincide may be biased, which may make it harder to count mismatches.

Overall, most research on motion and perception has been concerned with the perception of moving stimuli, rather than the perception of stimuli by a moving listener. To the best of our knowledge, no research has been concerned with factors influencing a moving listener's ability to perform crossmodal integration across the visual and auditory modalities. The field of ecological psychoacoustics is surprisingly quiet on this issue (Neuhoff, 2004).

Summary of Crossmodal Perception and Attention Studies

An analysis of the requirements for multimodal integration suggested by Newell (2004) indicates that two of the three necessary conditions for integration are achieved by the Mismatch task. In the Mismatch task, the visual and auditory information are both task-relevant and temporally congruent. The spatial congruence between the auditory and visual information, however, is different under some conditions. In the next, section I describe the mismatch task more fully and I draw from the applied and basic literature reviewed to provide a rationale for the specific hypotheses to be tested.

Experiment and Rationale

In the experiment to be described, participants will perform a continuous multimodal integration task in which they count the number of times the visual and auditory information about the behaviour of a set of objects does not match--the so-called “Mismatch” task. While they do the Mismatch task, participants will perform a secondary button-press task that will require them to walk around the test room and push labelled buttons on command. The Button-press task will create a dynamic visual background against which participants will have to perform the Mismatch task. In order to separate the influence of the Button-press task and the influence of walking on participants’ performance of the Mismatch task, the same button-press task will be used in a condition in which participants are sitting. The critical factor manipulated in the experiment is whether the auditory stimulus is delivered via speakers in the free-field or via a personal earpiece.

Rationale of Hypotheses

As shown in *Figure 3*, the first major hypothesis is that participants’ detection of mismatches will be more accurate when participants are sitting rather than walking, regardless of whether sound is delivered via free-field or earpiece. There are two reasons for this prediction. First, when participants are walking, the visual field behind the HMD display moves rather than remains stationary (Laramee & Ware, 2002). Participants who are walking may find it difficult to distinguish between soft and hard-looking shapes on the HMD, which will make it more difficult for them to detect mismatches. Second, walking absorbs attentional resources that can interfere with performance of a timeshared task (Sparrow et al., 2003). I expect participants to do worse at the Mismatch task when they are walking rather than sitting because of (1) structural competition between visual inputs (HMD display vs. visual guidance required for walking) and (2) general attentional resource competition between tasks.

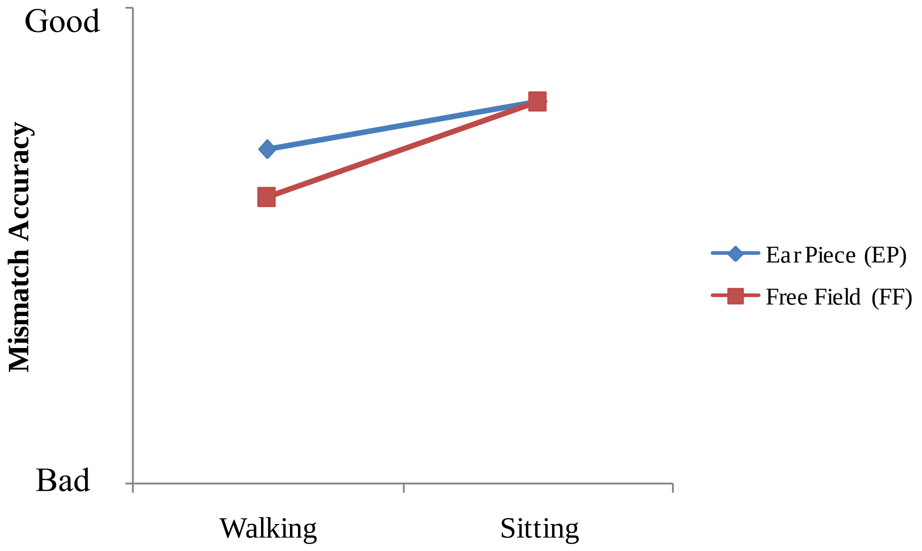


Figure 3. Predicted Mismatch Detection Accuracy

The second major hypothesis, as shown in *Figure 3*, is that there will be an interaction between movement (walking or sitting) and sound delivery (free-field or earpiece). First it is predicted that, when participants are sitting, there will be no difference in their ability to detect mismatches between the two methods of sound delivery. Second, when participants are walking, it will be more difficult for them to detect mismatches when sound is delivered in free-field rather than through an earpiece.

In the current experiment, participants are either walking or sitting and sound from the Mismatch task is delivered via free-field or earpiece. The relative spatial relationship between the sound and the vision on the HMD, therefore, may be different in some conditions. When participants are seated, the sound comes from a consistent direction with respect to the vision on the HMD, regardless of whether sound is delivered via earpiece or in free-field. When participants are walking, however, the direction from which the sound comes differs across the earpiece and free-field conditions with respect to the vision on the HMD. In particular, the relative spatial relationship between HMD vision and free-field sound changes dynamically when participants are walking. I conjecture that such changes will make crossmodal integration more difficult for participants walking and using free-field sound for two reasons (Newell, 2004). First, the azimuth of the sound with respect to the azimuth of the visual stimuli is usually well outside the range at which the visual stimulus captures the sound (Slutsky & Recanzone, 2001). Second, the azimuth of the sound continually changes (Lewald et al., 2001).

To reiterate, when participants are seated, the maintenance of spatial consistency will make integration of sound and vision equally likely with either sound delivery method. When participants are sitting, therefore, their mismatch detection accuracy will be equal whether sound is delivered via free-field or earpiece. When participants are walking, however, spatial consistency of sound and vision is only maintained with an earpiece and not with sound delivered in free-field. The inconsistent spatial relationship that exists between free-field sound and the vision on the HMD is less likely to satisfy the requirements for integration (Newell, 2004). When walking, therefore, participants' mismatch detection will be more accurate when sound is delivered via earpiece rather than free-field.

Hypotheses

The above conjectures are formalised as hypotheses below. *Figure 3* illustrates the general form of the predictions.

- H1: Participants will detect mismatches more accurately when they are sitting than when they are walking.
- H2: When participants are seated, they will detect mismatches equally well between the two methods of sound delivery (earpiece or free-field). When participants are walking, however, they will detect mismatches better with an earpiece than with free-field sound delivery.

Hypothesis 1 will be tested by seeking a main effect of Movement in an analysis of variance. Hypothesis 2 will be tested by seeking an interaction between Movement and

Sound Delivery in an analysis of variance, followed by a planned comparison between the Earpiece vs. Free-field methods of Sound Delivery at each level of Movement.

There are further minor hypotheses that follow from aspects of the two principal hypotheses above or from the intentions behind the experimental design. First, because the button press task has been designed to require people to walk and so create a dynamic visual background to the HMD display in the walking condition, rather than to generate competition for cognitive resources, participants' accuracy at pressing buttons should be equal across all conditions. Participants' latency at pressing buttons is expected to be longer when they are walking rather than sitting simply because of the extra time needed to walk to the next button. Second, participants' subjective workload for the walking conditions will be higher than for the sitting conditions because of the extra workload involved in walking.

Method

Contributions

The high-level motivation of the study was conceived by Penny Sanderson as well as the preparation of an Excel spreadsheet for data integration. Questionnaire design, experimental design, Mismatch task design, Button-press task design and counterbalancing were handled jointly by Penny Sanderson and me. Experiment technical set-up was carried out by Phil Cole and me. Mismatch software was coded by Mark Corben and Button-boxes were assembled by Nick Sibbald. The following responsibilities for the *Mismatch task* were handled solely by me: software requirement specifications, supervision of software coding, software testing, scenario design, sound design and shape design. The following work for the *Button-box task* was handled solely by me: hardware specifications, supervision of hardware development, software development in E-Prime and software testing. Finally, the following general *experiment tasks* were handled solely by MT: configuration of UQUL for video recording and digital capture, conducting mismatch sound pilot study, conducting experiment pilot study, conducting experiment, processing of data into summary scores for statistical analysis, and data collection, validation and reliability checks.

Participants

The study was approved by the School of Psychology ethics committee (approval number 07-PSYCH-4-121-JM). Participants were first-year psychology students from The University of Queensland who earned course credit for their participation. Participants were 11 males and 9 females (N = 20) aged between 17 and 39 (M = 22.3, SD = 5.6). All participants self-reported normal to corrected normal vision and normal hearing.

Design

The independent variables were Movement (Walking and Sitting) and method of Sound Delivery (Free-field and Earpiece) (see Table 1). The experiment used a repeated-measures design, so that all participants experienced all four of the conditions shown in Table 1 .

Table 1. 2 x 2 Repeated-Measures Experimental Design

		Sound Delivery Method	
		Free-field (FF)	Earpiece (EP)
Movement	Walking (Walk)	Walk-FF	Walk-EP
	Sitting (Sit)	Sit-FF	Sit-EP

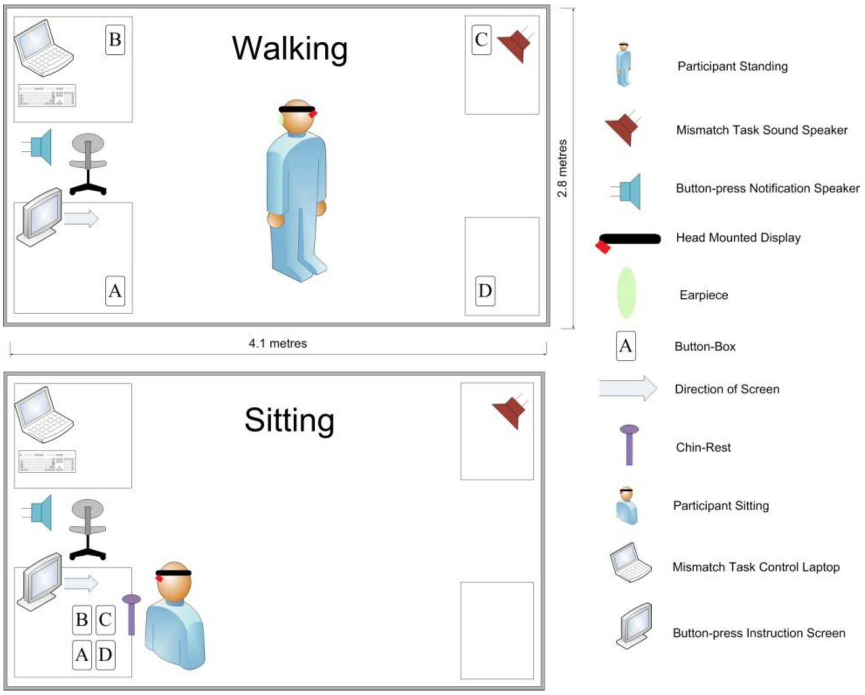


Figure 4. Experimental Layout for Walking and Sitting Conditions (Earpiece not Visible in Sitting Condition).

In the Walking condition, as can be seen in *Figure 4*, a button-box was placed at each corner of the test room so that participants had to move around the room in order to press the buttons. In the Sitting condition participants sat at a table with the button boxes in front of them with their head stabilised on a chin rest to limit motion of the HMD across the background (see *Figure 4* and *Figure 5*). The chin rest also served to limit the movement of sound relative to the participant.

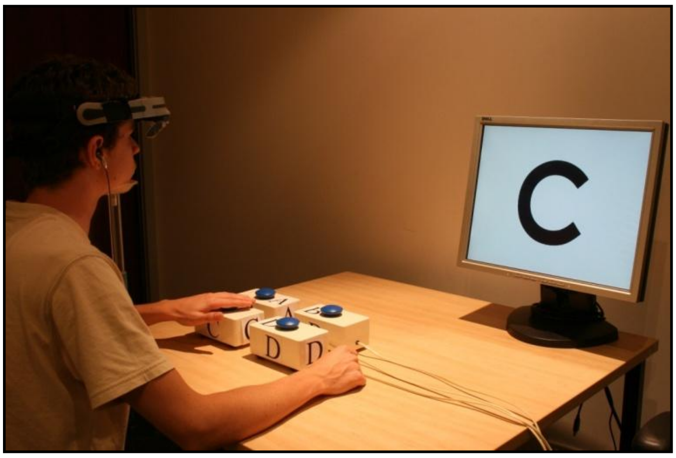


Figure 5. Participant in the Sitting Condition.

In the Free-field condition (FF), the sounds associated with object collisions came from a speaker in the corner of the test room. In the Earpiece condition (EP), the sounds associated with object collisions came from an earpiece which was placed in the participant's right ear (see *Figure 5*).

All participants experienced all four conditions created by crossing the above two variables. Order of presentation of conditions was counterbalanced using the Latin square method so that across each successive group of four participants, each condition was preceded and followed by all other conditions equally often, and each condition appeared at each serial position equally often. The order in which participants experienced the experimental conditions is shown in Table 2.

Table 2. Order of Experimental Conditions and Counterbalancing

Order	Condition			
1	Sit-EP	Sit -FF	Walk-FF	Walk -EP
2	Sit -FF	Walk -EP	Sit -EP	Walk -FF
3	Walk -EP	Walk -FF	Sit -FF	Sit -EP
4	Walk -FF	Sit -EP	Walk -EP	Sit -FF

The number of mismatches in each Mismatch task scenario was a controlled variable in the experimental design rather than an independent variable. In each experimental condition there were two separate trials of four minutes each. One trial used a scenario with a relatively high rate of mismatches and the other trial used a scenario with a relatively low rate of mismatches. Although the experimental conditions were counterbalanced so as to remove sequence or carryover effects, the Mismatch task scenarios were presented in the same sequential order across participants (see the ‘Crossmodal Object Integration Mismatch Task’ section for further details). Pilot testing established that the object bounce scenarios were sufficiently long and complex that the complexity added to the design by any attempt to counterbalance object bounce sequences as well as the experiment conditions could create opportunities procedural error. The counterbalancing ensured that each experimental condition was observed with all object bounce scenarios, and in all serial positions. Button-press scenarios were created dynamically at run-time, with the sequence of buttons determined by a strongly-constrained quasi-random sequence (see the ‘

Button-press Task’ section for further details).

The dependent variables were the accuracy with which participants counted the number of crossmodal object integration mismatches, the accuracy and latency with which participants performed the secondary Button-press task and participants' responses to questionnaires about the ease of performing the various tasks and the overall mental workload experienced. There were eight experimental trials in total with two trials for each of the four experimental conditions and a Post-condition questionnaire after each experimental condition. There was also a Post-experiment questionnaire.

Apparatus

Crossmodal Object Integration Mismatch Task

Description of mismatch task. The primary task was a crossmodal object integration mismatch task presented visually on the HMD and auditorily either via free-field speakers or via an earpiece. It will henceforth be referred to as the 'Mismatch' task. The participant's task was to determine whether the visual and auditory behaviour of the objects was congruent when the objects bounced against each other. The participant kept a silent mental count of the number of times the visual and auditory behaviour of the objects mismatched.

Figure 6 shows the visual display that was presented on the HMD. Three objects moved around the screen bouncing off each other and off the walls. There were two soft objects and one hard object. The soft objects had vaguely-defined edges suggesting they were covered in wool or another soft material. The hard object had sharp, well-defined edges suggesting it was made of wood or metal. The surrounding wall was defined as being hard for similar reasons.

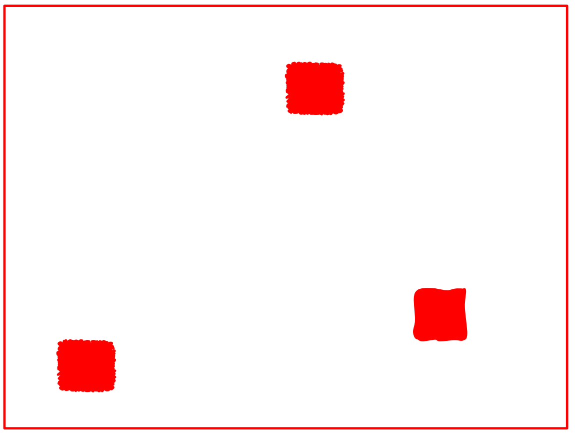


Figure 6. Objects moving on the Visual Display Shown on the HMD.

The sounds heard when objects collided were as follows:

- Soft object hits soft object → soft sound (correct)
- Hard object and soft object hit → soft sound (correct)
- Hard object hits wall → hard sound (correct)
- Soft object hits wall → soft sound (correct)

The Mismatch task involved detecting violations of the above behaviour, referred to as "mismatches". A pilot study revealed that people have strong and consistent associations with the set of shape and sound combinations that have been chosen. The sounds heard when mismatches occurred were the opposite of the correct sounds, as follows:

- Soft object hits soft object → hard sound (mismatch)
- Hard object and soft object hit → hard sound (mismatch)
- Hard object hits wall → soft sound (mismatch)
- Soft object hits wall → hard sound (mismatch)

Note that in this version of the Mismatch task there were never two hard objects moving on the screen.

Description of mismatch task scenarios. The starting position and initial vector of travel for the objects in the Mismatch task was different for each scenario. Eight unique scenarios were used for the experimental conditions. See Appendix A for the statistics of each scenario, including the number and type of collisions that occur.

The rate of mismatches was either High or Low. In Low mismatch rate conditions, the number of mismatches was 22, 23, 27 and 27 (7.4, 7.0, 8.6, and 8.6 percent of total bounces). In the High mismatch rate conditions the number of mismatches was 33, 34, 35 and 37 (10.3, 10.8, 10.9 and 11.3 percent of total bounces). In both cases, mismatches were timed so that the inter-mismatch-interval was always greater than three seconds, meaning that mismatches did not occur within three seconds of each other.

Button-press Task

Description of button-press task.

The purpose of the Button-press task was to ensure that participants moved dynamically around the room, experiencing the free-field sound from variable azimuth and viewing the HMD display against an often-dynamic background field. Four button boxes labelled A, B, C and D were placed one in each of the four corners of the room in the walking condition and were placed in a similar configuration in front of the participant in the sitting condition (see *Figure 4* and *Figure 5*). The button boxes were raised approximately 70cm off the ground so the buttons would be easy for the participant to press without bending down.

A large letter corresponding to the button box to press was presented on a computer screen at the front of the room. The letter changed every eight seconds. A notification sound was played when the letter changed so that participants did not have to constantly monitor the computer screen for the next letter but could glance at towards the screen when they heard the sound. The notification sound was selected to be quite different from the bounce sounds; it was a high-pitched, multiple-toned sound with a long attack and decay. The notification sound came from a speaker on the floor of the test room and to the right of the button-press letter screen (see *Figure 4*).

Letters were chosen randomly without replacement by the computer from the set {A, B, C, D} until all four letters had been selected. Then the set was restored and the random selection without replacement was repeated, so that no more than two letters would be presented successively. This procedure was repeated until 30 letters had been selected, resulting in a total of 30 button-press commands for each scenario. Because the letters were randomly generated within the constraints described above, the order of the letters was different for each scenario and for each participant. Accuracy and reaction time measures were collected for each button-press.

Description of Software and Hardware

Custom software was developed in C++ to create and run the Mismatch task. The software was run on a laptop computer (Pentium 4 2.66 GHz processor, 448 MB RAM, Windows XP Professional 2002 SP2) and a Sony VGN-U50 (900MHz Celeron-M processor, 256 MB RAM, Windows XP Professional SP2, 800 x 600 resolution) that were networked using and wireless network hub. The Mismatch task scenarios were selected and run using the laptop computer which controlled the shapes on the Sony U-50. The objects of the Mismatch task were approximately 80 x 80 pixels in size and travelled around the screen at approximately 150 pixels per second. A programming constraint exists where, very occasionally, and object collides with another object but bounces through it rather than off it and participants were told to treat the event as a usual collision.

The HMD used was a Nomad ND2000 Augmented Vision System with a single transparent monocle (resolution: 800 x 600, brightness: 303, 50%, ABC-on, contrast: 80-80%, Gamma: 5, video sample frequency: 1059-25%) and was connected to the Sony U-50 via VGA cable. Sounds generated by the Mismatch task from the Sony U-50 were sent to a loudspeaker (Edirol MA-7A Stereo Micro Monitor) using a wireless transmitter (Sony UHF Synthesized Diversity Tuner URX-P1) in the Free-field condition. The sound pressure level of the sounds, measured from the centre of the experiment room, was 70dB(A) max. In the Earpiece condition, the stereo ear buds (Sony MDR-E829V, with volume control) were connected to the Sony U-50.

The button-boxes used for the Button-press task were custom designed for this experiment. The boxes were (14.5cm long x 8cm wide x 6.5cm deep) and the button was a blue circular momentary push-button switch with a 4.5cm flat top (ITW switches 76-9440) that is easy to locate and press with minimal visual guidance.

The Button-press task was controlled with the E-Prime software (version 1.1.4.1). There was one block list and one trial list. The trial list was made up of four levels (A, B, C and D). Button-box input was sent to E-Prime via a games port and only one response was accepted per trial. The letter, in 280 point Century Gothic font, and the notification sound were displayed on an instruction slide. E-Prime was run on a desktop computer with a 17-inch LCD display and a resolution of 1280 x 1024 (Pentium 4 1.90 GHz processor, 512 MB RAM, Windows XP Professional 2002 SP2). The sound pressure level of the notification sound from the loudspeaker (Harmon/Kardon H/K 695-04) was approximately 64dB(A) max.

The experiment was recorded with four ceiling-mounted cameras in order to provide a check on the standardisation of experimental procedures and to capture participants' movements in the walking conditions for possible analysis in the future. The camera output was fed into a quad layout on a single screen and was captured in MPEG format on a Mac PowerBook (see *Figure 7*). Video analysis was not used for the current experiment.

Figure 7. Quad Video Layout with Participant in Walking Condition

Questionnaires

Participants filled out a background questionnaire before the experiment began, Post-condition questionnaires after both trials for each condition had been run, and Post-experiment questionnaires after all conditions had been run. The Eysenck Personality Questionnaire was administered at the very end of the experiment. The questionnaires are described in more detail below and are also included in the Appendices.

Background. The initial background questionnaire, shown in Appendix B, asked participants their age, gender, any hearing impediments or visual impediments, and it recorded the results of the Miles procedure for determining eye dominance (Miles, 1930).

Post-condition. The purpose of administering the Post-condition questionnaire was to collect information from participants about the condition they had just experienced while it was fresh in memory. The Post-condition questionnaire was administered after the last training trial to give participants practice with filling it out. Then it was administered when the participant completed each of the four experimental conditions. The questionnaire asked participants how easy it was (1) to detect the mismatches, (2) to keep count of the mismatches, (3) to integrate the vision with the sounds and (4) to timeshare the two tasks. Participants circled a number on a seven-point Likert scale and could make written comments or clarifications if they wished (see Appendix C).

Post-experiment. The purpose of administering the Post-Experiment questionnaire was to collect information from participants about their view of the relative ease of the four experimental conditions, after completing them all. The Post-experiment questionnaire contained the same questions as the Post-condition questionnaire but the questions were presented in a matrix format (see Appendix D). The arrangement helped participants to rate how easy each of the conditions was relative to the other conditions. Participants could also make written comments or clarifications.

In addition, the Post-Experiment questionnaire also asked participants how annoying they found the button box press reminder sound and the bounce sounds. Two open-ended questions probed participants' strategy for performing the tasks and a final question asked for any final comments about the experiment.

NASA-TLX. The purpose of administering the National Aeronautics and Space Administration Task Load Index (NASA-TLX) was to probe specific sources of workload that the participants might experience in the task. The NASA-TLX is a subjective workload measure assessing mental, physical and temporal demand as well as effort, performance and frustration levels. It was administered after participants completed each of the four experimental conditions. . Using the standardised 21-point scale layout of the NASA-TLX response sheet, participants circled a marker that best reflected their experience in the preceding condition (see Appendix E). In recognition of the arguments of Nygren (1991), the procedure for ranking subjective importance of the TLX subscales was not used.

Eysenck personality questionnaire. The Eysenck Personality Questionnaire (EPQ) was administered as part of a long-term strategy by the Cognitive Engineering Research Group to investigate the relationship between personality traits and perception of ease of use and tolerance of different kinds of advanced displays. The EPQ contains 90 forced-choice ‘yes’ or ‘no’ questions assessing traits of extraversion, psychoticism and neuroticism as well as social-desirability bias (Eysenck, 1980).

Procedure

The following shows the order in which the various treatments in the experiment were administered. More information about each phase of the procedure is given in the sections below.

- Formalities
  - Information Sheet
  - Informed Consent
  - Background Questionnaire
- Mismatch Task Training on Screen
  - No Mismatches
  - Explaining and demonstrating mismatches
  - Complete trial with mismatches
- Set-up
  - Earpiece volume
  - Fit HMD and belt
  - HMD focus
- Dual-task Training
  - Button-press training with no primary task
  - Sitting with earpiece
  - Walking with earpiece
- Condition 1
  - Trial 1
  - Trial 2
  - Post-condition Questionnaire
- Condition 2
  - Trial 3
  - Trial 4
  - Post-condition Questionnaire
- Condition 3
  - Trial 5
  - Trial 6
  - Post-condition Questionnaire
- Condition 4
  - Trial 7
  - Trial 8
  - Post-condition Questionnaire
- Post-experiment Questionnaire
- EPQ
- Debriefing

Formalities

Participants read the information sheet which provided a brief overview of the experiment and the tasks required . They then signed the informed consent form. Participants answered the background questionnaire and the experimenter determined the participant's dominant eye using the Miles procedure. Participants were told briefly about the two tasks that would be required of them and that there would be an initial training period.

Mismatch Task Training on LCD Screen

The Mismatch task was presented on the small LCD screen of the Sony U-50 in front of the participant while they were seated. This was so the participant could be trained on the fundamentals of the Mismatch task without the added complexity of wearing the HMD. For the first presentation of the Mismatch task, participants saw and heard the shapes moving but no mismatches occurred. While watching the objects moving on the screen, the experimenter explained the typical properties of the objects and the walls.

Once the participant was familiar with the fundamentals of the Mismatch task, a second presentation of the Mismatch task was started, now with mismatches occurring. The experimenter explained that occasionally a mismatch would occur between what is seen and what is heard – in other words, the visual and auditory object behaviour would not match. For training purposes the participant was asked to report to the experimenter any mismatches they observed until it was clear that they understood the task. A complete scenario with mismatches was then run, during which the participant was asked to maintain a silent mental count of the number of mismatches they observed and to report the total to the experimenter at the end of the scenario.

Set-up

Earpiece volume. The earpiece volume was adjusted by the participant until there was approximately equal subjective loudness between the speaker volume and the earpiece volume. The participant sat on a chair in the middle of the room facing the speaker, at approximately 180 cm from the speaker. The sound from the HMD Mismatch task was periodically switched between the speaker and the ear piece while the participant adjusted the personal volume control on their earpiece. The procedure continued until the free-field speaker sound and the earpiece sound were at the same subjective loudness. The volume level of the earpiece was then secured physically so it could not be altered during the experiment.

HMD adjustment and focus. The experimenter fitted the HMD and a belt with the Sony U50 and wireless technology around the participant's waist so the participant could move unimpeded around the test room. The experimenter then explained how the focus worked on the HMD. Letters of different size, similar to a Snellen eye chart, were presented on the HMD. Participants were asked to stand 130cm away from a wall of the test room and to keep their eyes focused on fine spatial detail of the test room blind. The participant then was encouraged to move the focus of the HMD to better experience the differences in focus and the motion of the focus dial.

The HMD focus was then placed at infinity and participants were asked to adjust the focus so that both the detail of the test-room blind and the Snellen-like letter array on the HMD screen were in focus without the participant having to adjust the accommodation of their eye. Participants were instructed on the importance of maintaining their visual focus on the test-room blind and using their peripheral vision to monitor the focus of the HMD screen until both were in focus. The focus setting on the HMD did not change during the experiment.

Dual-task Training

Participants were then trained to combine visual monitoring of the objects on the HMD and the Button-press task. The Button-press started and the shapes were presented on the HMD, but the Mismatch task sound was not played so the Mismatch task could not be done.

This helped participants become accustomed to listening for the Button-press notification sound and doing the Button-press task while watching the objects move on the HMD. Participants were told that the letter indicating the next button to press would change only when the Button-press notification sound was heard. They were asked to face each button box when they pressed its button. Participants were asked to do the Button-press task as efficiently as possible but that they should consider the Mismatch task as their primary and most important task.

Participants then completed one full training scenario in the sitting earpiece condition followed by a full training scenario in the walking free-field condition, using the Mismatch task sounds. They then filled in a Post-condition questionnaire with respect to their experience in the most recent walking-free-field condition.

Begin Conditions

Participants were then told they had completed training and that the experiment would now begin. They were informed that they would experience eight trials under different conditions. There would always be two trials in a row in the same condition. Each condition would be followed by the Post-condition questionnaire. Participants could take a break at any point during the experiment.

After all trials were completed, the Post-experiment questionnaire and the Eysenck Personality Questionnaire were administered. Finally there was a debriefing session in which the experimenter outlined the aims of the study and answered any questions.

Results

Details of the participant sample will be given. Then, results will be presented for dependent variables that were critical to the major hypotheses, including mismatch accuracy and the button-press data. Finally, the results of other potentially informative dependent variables will be reported, including post-condition questionnaire, post-experiment questionnaire, NASA-TLX and Eysenck Personality questionnaire.

Participant Sample

Two participants could not complete the study because the earpiece size was not appropriate for their ear canal. One participant reported one mismatch count that was an outlier (> 2 *SD* above the mean) and they reported high levels of fatigue and low motivation. The participant's data was therefore replaced with a further participant's data using the same counter-balancing sequence.

Mismatch Accuracy Data

A 2 x 2 x 2 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, $F(1,19) = 8.895$, $MSE = 0.021$, $p = 0.008$, *partial* $\eta^2 = 0.319$. When they were sitting ($M = 83\%$, $SD = 15.6\%$), participants counted mismatches more accurately than when they were walking ($M = 76\%$, $SD = 18.7\%$). There was no main effect of Sound Delivery. There was, however, a significant main effect of Mismatch Rate, $F(1,19)=78.110$, $MSE = 0.014$, $p < 0.001$, *partial* $\eta^2 = 0.804$. Participants counted mismatches more accurately when the Mismatch Rate was low than when it was high.

As predicted, there was a significant two-way interaction between Movement and Sound Delivery, $F(1,19) = 6.440$, $MSE = 0.023$, $p = 0.020$, *partial* $\eta^2 = 0.055$, which is shown in

Figure 8. However, the results of planned comparisons comparing Earpiece vs. Free-field at each level of Movement were not as hypothesised. Specifically, when participants were sitting, they counted mismatches more accurately with sound delivered in free-field (*M* = 87.6%, *SD* = 17.5%) than with sound delivered through an earpiece (*M* = 79.4%, *SD* = 16.9%), *p* = 0.020. When walking, however, participants counted mismatches equally well with sound delivered in either free-field (*M* = 74.7%, *SD* = 19.6%) or earpiece (*M* = 78.7%, *SD* = 20.0%), *ns*. A post-hoc Tukey HSD test was used to interpret the interaction shown in *Figure 8*. When listening to the sound in free-field, participants counted mismatches more accurately when they were sitting than when they were walking, *p* = 0.006. No other specific post-hoc contrasts were significant (see Table F2).

There was also a significant two-way interaction between Sound Delivery and Mismatch Rate, $F(1,19) = 5.124$, $MSE = 0.008$, $p = 0.035$, *partial* $\eta^2 = 0.213$. The two-way interaction between Movement and Mismatch Rate was non-significant and the three-way interaction between Movement, Sound Delivery and Mismatch Rate was not significant (see Table F1 for variance table).

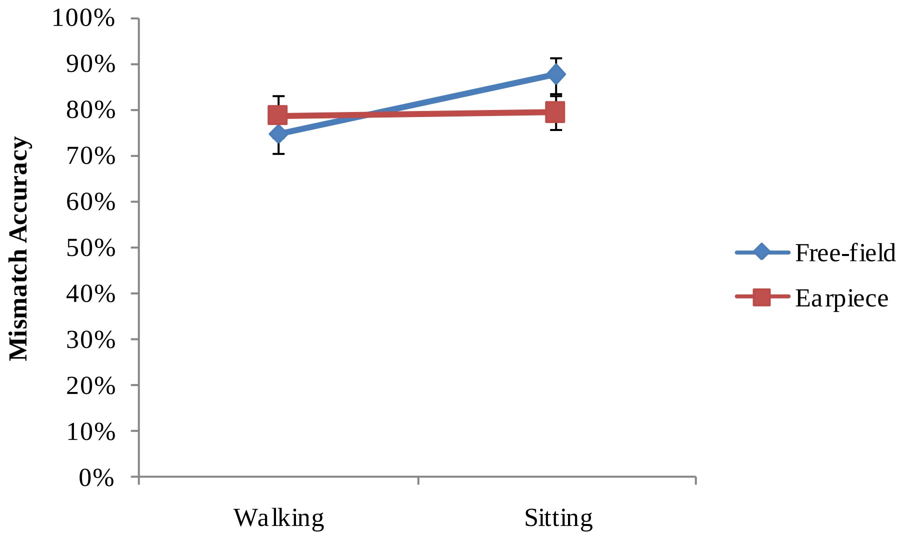


Figure 8. Mismatch Detection Accuracy (+ MSE) for Movement (Walking vs. Sitting) and Sound Delivery (Free-field vs. Earpiece) Conditions.

Button-press Performance

A 2 x 2 x 2 x 4 repeated measures ANOVA was conducted on the button-press reaction time data with the within-subjects factors of Movement (Sitting vs. Walking), Sound Delivery (Free-field vs. Earpiece), Mismatch Rate (Low vs. High) and Button Identity (A vs. B vs. C vs. D). Results are shown in Figure 9 without the Sound Delivery factor (see Table G1 for variance table).

As predicted, there was a significant main effect of Movement in how quickly participants pressed buttons, $F(1,19) = 346.517$, $MSE < 0.001$, $p < 0.001$, *partial* $\eta^2 = 0.948$. When they were sitting, participants pressed buttons faster than when they were walking. Unexpectedly, there was also a significant main effect of Button Identity, $F(3,57) = 3.116$, $MSE < 0.001$, $p = 0.033$, *partial* $\eta^2 = 0.141$. A post-hoc Tukey HSD test indicated that participants' reaction time to button A was faster than to button C, $p = 0.028$. No other specific post-hoc contrasts were significant (see Table G2). All other main effects and interactions, for button-press reaction time, were non-significant (see Table G1).

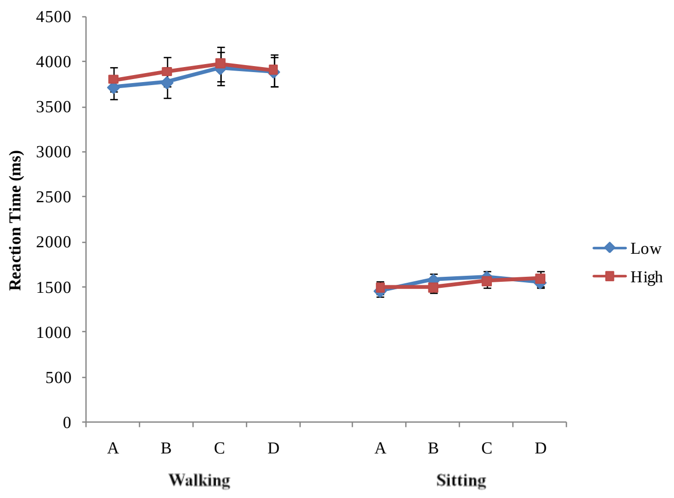


Figure 9. Reaction Times to Button-press Task for the Four Buttons with High and Low Mismatch Rate when Walking or Sitting.

As noted, the Button-press task was intended to be performed equally accurately in all conditions. Button-press accuracy ranged from 99.3% correct to 99.8% correct across conditions. Because of the extremely accurate levels of performance in all conditions, the data satisfied the assumptions neither of parametric tests (for example, ANOVA) nor of non-parametric tests (for example, Friedman's test of signed rank for correlated samples). Therefore no inferential tests were run on the data.

Questionnaires

Post-Condition and Post-experiment

Results for participants' post-condition and post-experiment responses to each question will be presented in turn. In each case, a 2 x 2 repeated measures ANOVA was conducted with the within-subjects factors of Movement (Sitting vs. Walking) and Sound Delivery (Free-field vs. Earpiece). Means and standard deviations for questionnaire responses are omitted to aid readability. Instead, Figure 10 shows the results from both questionnaires for side-by-side comparison. As will be seen, participants responded somewhat differently when making ratings after each condition than when making ratings at the end of the experiment.

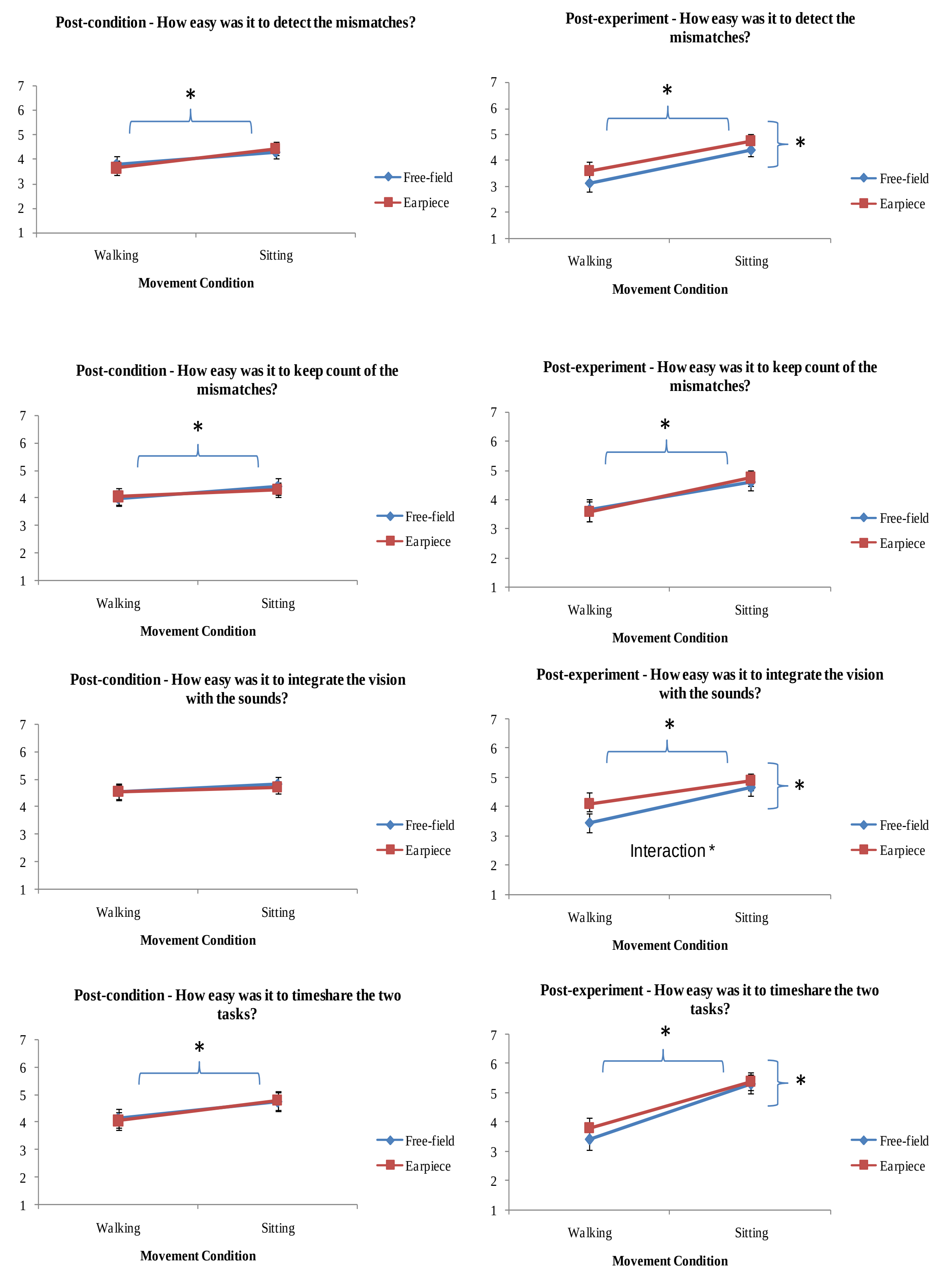


Figure 10. Post-condition and Post-experiment Questionnaire Results

How easy was it to detect the mismatches?

When participants rated after each condition how easy it was to detect mismatches, there was a significant main effect of Movement, $F(1,18) = 18.853$, $MSE = 0.509$, $p < 0.001$, *partial* $\eta^2 = 0.512$. Neither the main effect of Sound Delivery nor the interaction between Movement and Sound Delivery was significant (see Table I1).

When participants rated at the end of the experiment how easy it was to detect mismatches, there were significant main effects of both Movement, $F(1,18) = 20.886$, $MSE = 1.379$, $p < 0.001$, *partial* $\eta^2 = 0.521$, and Sound Delivery, $F(1,18) = 5.630$, $MSE = 0.568$, $p = 0.028$, *partial* $\eta^2 = 0.229$. The interaction between Movement and Sound delivery was non-significant (see Table I2).

In summary, participants always reported that it was easier to detect mismatches when sitting than when walking. Only after at the end of the experiment, however, did they rate mismatch detection easier with an earpiece than with free-field sound.

How easy was it to keep count of the mismatches?

When participants rated after each condition how easy it was to keep count of mismatches, there was a significant main effect of Movement, $F(1,18) = 4.579$, $MSE = 0.830$, $p = 0.046$, *partial* $\eta^2 = 0.201$. After each condition, participants reported that it was easier to keep count of mismatches when sitting than when walking. Neither the main effect of Sound Delivery nor the interaction between Movement and Sound Delivery was significant (see table I3).

When participants rated at the end of the experiment how easy it was to detect mismatches, there was a significant main effect of Movement, $F(1,18) = 9.291$, $MSE = 1.555$, $p = 0.007$, *partial* $\eta^2 = 0.321$. Participants reported that it was easier to detect mismatches when sitting than when walking. Neither the interaction main effect of Sound Delivery nor the interaction between Movement and Sound Delivery was significant (see Table I4).

How easy was it to integrate the visual information with the sounds?

When participants rated after each condition how easy it was to integrate the visual information with the sounds, neither the main effects of Movement nor Sound Delivery, nor the interaction between the two reached significance. Immediately after finishing each condition, participants reported that it was equally easy to integrate the visual information with the sounds (see Table I5).

When participants rated at the end of the experiment how easy it was to integrate the visual information with the sounds, there were significant main effects of Movement, $F(1,18) = 13.333$, $MSE = 1.500$, $p = 0.002$, *partial* $\eta^2 = 0.412$, and Sound Delivery, and $F(1,18) = 7.364$, $MSE = 4.050$, $p = 0.014$, *partial* $\eta^2 = 0.279$. Participants reported that it was easier to integrate the visual information with the sounds when walking than when sitting and with an earpiece than with free-field sound. There was also a significant two-way interaction between Movement and Sound Delivery, $F(1,19) = 5.630$, $MSE = 0.142$, $p = 0.028$, *partial* $\eta^2 = 0.229$ (see Table I6).

How easy was it to timeshare monitoring for mismatches and pushing the buttons?

When participants rated after each condition how easy participants found it to timeshare monitoring, there was a significant main effect of Movement, $F(1,18) = 10.365$, $MSE = 0.925$, $p = 0.005$, *partial* $\eta^2 = 0.365$. Participants reported that it was easier to timeshare when sitting than when walking. Neither the main effect of Sound Delivery nor the interaction between Movement and Sound Delivery was significant (see Table I7).

When participants rated at the end of the experiment how easy it was to timeshare, there was a significant main effect of both Movement, $F(1,18) =27.544$, $MSE=2.224$, $p < 0.001$, *partial* $\eta^2 = 0.591$, and Sound Delivery, $F(1,18) = 4.524$, $MSE = 0.276$, $p = 0.047$, *partial* $\eta^2 = 0.192$. Participants reported that it was easier to timeshare when sitting than when walking and easier with an earpiece than with free-field sound. The interaction between Movement and Sound Delivery was non-significant (see Table I8).

NASA-TLX

A 2 x 2 x 6 repeated measures ANOVA was conducted on participants' responses to the NASA-Task Load Index. The within-subjects factors were Movement (Sitting vs. Walking), Sound Delivery (Free-field vs. Earpiece), and TLX Subscale (Mental Demand vs. Physical Demand vs. Temporal Demand vs. Effort vs. Performance vs. Frustration) Results are shown in **Error! Reference source not found.** and ANOVA results are shown in Table J1.

There was a significant main effect of Movement in how participants rated their workload, $F(1,18) = 45.805$, $MSE = 11.450$, $p < 0.001$, *partial* $\eta^2 = 0.718$. Participants' subjective workload was higher when they were walking than when they were sitting. There also was a significant main effect of TLX-Subscale, $F(5,90) = 16.123$, $MSE = 39.35$, $p < 0.001$, *partial* $\eta^2 = 0.473$, with Mental Demand rated highest and Physical Demand lowest in workload intensity. A significant interaction was found between Movement and TLX-Subscale, $F(5,90) = 2.769$, $MSE = 7.78$, $p = 0.023$, *partial* $\eta^2 = 0.133$. A Tukey HSD revealed that Physical Demand was the only subscale for which participants rating their workload significantly higher when they were walking than when they were sitting (see Table J2).

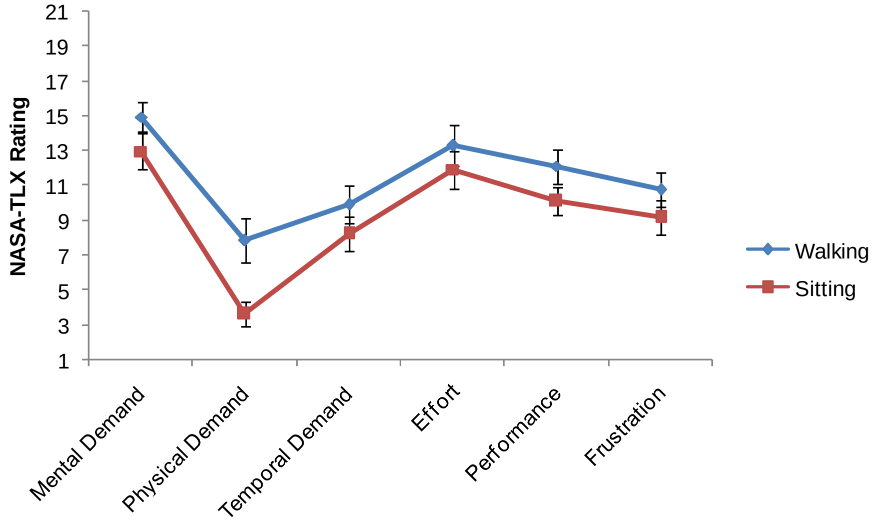


Figure 11. NASA-TLX Subjective Workload Ratings

Findings with the EPQ

As noted earlier, all participants completed the EPQ as part of an ongoing laboratory investigation into potential relationships between participants' tolerance of advanced displays and personality type. The analysis is exploratory and not a principal focus of the thesis, so the results are just reported briefly here.

Participants' scores on the four subscales of the EPQ – psychoticism, extraversion, neuroticism and lie scale – were correlated with participants' mismatch count data, post-condition questionnaire results, and post-experiment questionnaire results using Pearson product-moment correlation coefficients. In the interests of finding patterns of results, Bonferroni corrections were not applied. The results are shown in Appendix K.

All but one of the 10 significant correlations in Appendix K is with the Psychoticism subscale. The negative correlations indicate that low levels of Psychoticism (which is usually interpreted as high levels of empathy) are associated with better performance on the Mismatch task and higher ratings of ease in the questionnaires. Eight of the nine correlations with questionnaires relate specifically to ease of detecting mismatches or ease of integrating the visual and auditory information. Eight of the 10 correlations overall are between Psychoticism and either performance when sitting or ratings of detection ease in the sitting conditions. Performance on the Mismatch task when sitting with an earpiece was significantly negatively correlated with Psychoticism.

Discussion

The aim of this study was to determine whether there is cause for concern about the performance of people using Head-Mounted Displays (HMDs) and auditory displays together under particular conditions. The effect of movement and sound delivery on people's ability to perform a novel crossmodal mismatch detection task was investigated. The results of the study will now be reviewed and their theoretical and practical implications will be outlined. Limitations will be identified and future research will be proposed.

In this study, two factors were considered that reflect the multimodal nature of dynamic work environments –movement (sitting or walking) and the sound delivery method (free-field or earpiece). A task was created where participants had to integrate auditory and visual information to detect mismatches in object properties. The first hypothesis was that participants would count mismatches better when they were sitting than when they were walking. This hypothesis was supported by the data. However, this result is only meaningful when considered in the context of the interaction between participant movement and sound delivery.

The second hypothesis was that there would be an interaction such that; when participants were sitting they would count mismatches equally well, but when walking they would count mismatches more accurately when the sound came from the earpiece than from free-field. This hypothesis was not supported by the data, although there was a significant interaction. To aid interpretation, the interaction as a whole was investigated by considering all possible contrasts. With sound in an earpiece, participants counted mismatches equally well whether they were sitting or walking (both conditions 79%). With sound in the free-field, however, participants counted mismatches more accurately when they were sitting (87%) than when they were walking (74%). It appears that when people integrate sound from an earpiece and vision from an HMD, it makes no difference to their performance whether they are sitting or walking. However, when people integrate sound from the free-field and vision from an HMD, their performance is worse when walking than when sitting.

Findings from minor hypotheses include the Button-press task, questionnaires and subjective workload responses. The Button-press task was designed as a means to motivate movement in the walking condition and control for workload in the sitting condition. It appears to have been effective for this purpose because results show very high accuracy overall and no difference in accuracy across any condition, suggesting that it served its function of creating movement rather than mental workload. As predicted, NASA-TLX workload ratings were higher when participants were walking rather than sitting. Ratings from the physical workload scale when walking, however, were approximately two and a half times larger than the average of the other subscales. As expected, participants took longer to press buttons when they were walking than when they were sitting, largely because it took time to walk to the buttons. Overall, it appears the Button-box task did not generate any competition for attentional resources.

There were systematic differences in how participants responded to the post-condition versus post-experiment questions about ease of detecting mismatches, ease of counting mismatches, ease of integrating the sound and vision, and ease of timesharing. In both questionnaires, participants generally reported that the task was harder when they were walking rather than sitting. Only when asked post-experiment, however, did participants report that the task was easier with an earpiece rather than free-field. The theoretical and practical implications of the present study's results will now be discussed.

Theoretical Implications

The first hypothesis predicted a main effect of movement such that participants would detect mismatches more accurately when sitting rather than walking. This was expected on the basis that: (1) visual interference from a moving background behind the HMD would make it difficult to distinguish soft and hard-looking shapes (Laramee & Ware, 2002) and (2) attentional resources absorbed by walking (rather than the Button-box task itself) would interfere with the Mismatch task (Sparrow et al., 2003). The current findings suggest, surprisingly, that walking had no effect people's ability to detect mismatches when they use an earpiece. This finding conflicts with those of Laramee and Ware (2002) who show that a moving image behind an HMD leads to worse performance on an HMD task. It also contrasts with Sparrow et al. (2003) who show that walking absorbs attentional resources, such that people perform worse at secondary tasks. It is difficult to reconcile prior literature with the results of the current study. It appears that walking does not influence people's ability to integrate visual and auditory information when they have an earpiece.

The current findings suggest that that the performance cost observed, when people are walking with free-field sound, is not due to their walking. Walking, in this case, is not a sufficient explanation for the performance cost as the same cost should be observed when people are walking with an earpiece. An alternative explanation is that the inconsistent spatial relationship between the vision on the HMD (fixed to the head) and the sound from free-field (fixed to the room), made it more difficult for people to give an accurate mismatch count. This inconsistent spatial relationship may make it more difficult for people to integrate auditory and visual stimuli.

The Mismatch task satisfied the task-relevant and temporally congruent conditions necessary for crossmodal integration in all conditions. Spatial consistency, however, between visual and auditory stimuli was particularly different when walking with free-field sound compared to the other conditions. Within certain spatial and temporal limits, vision and sound will combine to create a unified percept of having come from the same object (Lewald, et al., 2001; Newell, 2004).

Early ventriloquism research suggested that an auditory stimulus will be “captured” by an available visual stimulus but not if the auditory stimulus occurs more than 300 msec from the visual stimulus (Slutsky & Recanzone, 2001). In addition, the more spatially separated the auditory and visual stimuli are, the less likely it is that the auditory stimulus will be captured by the visual stimulus (Slutsky & Recanzone, 2001). Subsequently, researchers such as Ernst and Banks (2002) and Alais and Burr (2004) have shown that the degree of “capture” is a function, not just of displacement, but also of the uncertainty of stimulus location. Whether the visual stimuli and sound are moving in the same or different directions makes no difference – only the relative uncertainties of stimulus location makes a difference (Alais and Burr, 2004).

In the current study, two factors may be contributing to mismatch counts being less accurate while walking with free-field sound: (1) walking may create greater uncertainty in the source location of sound and (2) walking with free-field sound creates greater spatial separations between HMD azimuth and sound azimuth. The former factor suggests that the sound is not more likely to be captured by the vision. In the walking condition, the mean orientation of free-field sound to HMD vision is mostly outside the range at which object integration will occur (Jack & Thurlow, 1973; Thurlow & Jack, 1973). The latter factor would make the vision and sound associated with bounces less likely to be perceived as an integrated crossmodal event, and mismatches therefore harder to detect. It may be that the dynamic changes in sound created when the participant walks around are not matched by dynamic changes in the visual stimulus, creating a challenge for object integration and leading to worse mismatch count accuracy. When the participant wears an earpiece, however, the sound has consistent acoustic properties and lies at a consistent azimuth from the HMD display, which might lessen the challenges to object integration and lead to better mismatch count accuracy.

Despite the above speculations, it is difficult to provide a clear cut explanation for the findings of the current study using prior literature. As with the experimental hypotheses, the logical deductions made to explain the results are tenuous. This is largely because most prior research on motion and auditory stimuli has been concerned with the perception of moving auditory stimuli, rather than the perception of auditory stimuli by a moving listener (Calvert et al., 2004). To the best of my knowledge, no research has investigated the factors influencing a moving listener's ability to perform crossmodal integration across the visual and auditory modalities. Despite this, humans commonly perform crossmodal integration as we move through the world. The current thesis, therefore, appears to have opened up an area of potential research that is highly representative of real world situations but that is not illuminated by current theories of crossmodal integration (Calvert et al., 2004) or of ecological psychoacoustics (Neuhoff, 2004).

Practical Implications

Despite the limited scope of the study, I propose that there may be a cause for concern for people using both HMDs and auditory displays in dynamic environments. First, there may be performance costs, associated with an audio-visual integration task, with free-field sound when people are walking compared to free-field sound when people are sitting. Contrary to predictions, when walking, people were not better at performing the mismatch task with the earpiece than with free-field sound. It may be possible to raise people's performance with head-mounted audio by using headphones – this is described in the following section. If a recommendation had to be made based on what is currently known, delivering sound through an earpiece may be the best option. Although performance with an earpiece may not be as high as other potential sound delivery methods, it is not sensitive to movement and is not reacting to a dynamic situation.

Another interesting finding is that people's beliefs about their own performance was not consistent with their actual performance. This phenomenon is similar to that obtained by Kallinen and Ravaja (2007), as mentioned previously, when they compared headphone vs. speaker sound delivery. The present results also suggest that the extra difficulty people reported for walking was not reflected in their walking performance on the Mismatch task with an earpiece. People may have felt that walking made the task more difficult but they tried harder to compensate for this. Further, people generally believed that their performance was better with an earpiece when walking than when sitting, where-as the behavioural data suggest that this is not the case. The possibility of demand characteristics, however, cannot be discounted. It may be that people interpreted the purpose of the experiment and responded to the questionnaires accordingly. This is unlikely though, as it is not obvious what the demands are.

Findings from these questionnaire data have may implications for the issue of user acceptance of advanced technologies (Dillon & Morris, 1996). The present results show that people's beliefs about ability to perform the task under different conditions did not match their actually ability to perform the task. This highlights the importance of assessing people's actual performance with advanced displays rather than their subjective opinions. This is especially important in safety-critical environments where safe and effective performance is of greater importance than subjective attitudes toward technology. People's resistance to advanced technologies may be offset by demonstrating to people the conditions under which their performance will be better.

Use of the EPQ in this study is exploratory and the results are speculative. There seems to be a relationship between how easy people say the task is and how empathic they are. It may be that people who are very empathic tend to make more positive responses about the ease of the task. Sanderson et al. (2007), however, found a correlation between people's empathy and reporting that patient monitoring was easy only when auditory displays were included, but not when visual displays were used alone. Although an auditory display was always used in the current experiment, present results are at least consistent with this prior finding. The following assertion is highly speculative but perhaps empathic people are more likely to welcome a personal auditory display.

Limitations and Future Research

Seven limitations of the present study will now be discussed and a suggestion of future research will be made in some cases. First, it is important to note that the results of this study are specific to transparent monocular HMDs. An opaque HMD would not be subject to visual interference when people are walking. Second, only a task that requires crossmodal integration of sound and vision has been considered. The effect of sound delivery method and movement on tasks that can be performed with vision or sound alone (‘redundant’ tasks) is unknown. Third, the Button-press task presented both visual and auditory information that was separate from the Mismatch task. The extra task-relevant stimuli may have affected crossmodal integration in the Mismatch task. Fourth, only a dynamic display was presented on the HMD. How a static visual display (e.g. Laramee & Ware, 2002) might affect audio-visual integration is unknown. Fifth, the current experimental set-up makes it difficult to match the properties of the sound stimuli across earpiece and speaker, which makes direct comparison problematic. Sixth, the apparent location of the auditory display was at people’s right side with an earpiece in the right ear and may not have been the best match for the forward location of the visual display.

Finally, an important theoretical question raised by this research is whether the Mismatch task actually requires low-level ‘perceptual’ integration of sound and vision or whether the combination of the sound and vision occurs during high-level processes. As the Mismatch task is novel and has only been tested in one experimental situation, its reliability and validity as an indicator of crossmodal integration is unknown. Basic research relating to when a crossmodal integration task is classified as low or high-level is currently a matter of debate (Alais, Morrone, & Burr, 2006; Driver & Spence, 1998). Alais, Morrone and Burr suggest that attentional resources for crossmodal integration may be employed in different ways depending on the nature of the task demands. Despite these concerns, it is clear that the mismatch task is sensitive to some extent.

As previously mentioned, an earpiece may not have produced an apparent location for the sound that was the best match for the apparent location of the visual display and, therefore, not produced best crossmodal integration. The visual display was in front of the participants and the auditory display to the right with an earpiece in the right ear. A consistent audio-visual relationship seems to help but having the auditory and visual information in the exact same location may make performance even better (Driver & Spence, 1994). A different head-mounted auditory delivery system might make sound vision from the HMD appear to be spatially co-located. A follow-up study will compare headphone vs. free-field delivery. Having sound in both ears, from a set of headphones or two earpieces, may place the perceived location of the sound source in the centre of the head. This arrangement may make it easier for people to integrate auditory and visual information to form the perception of a single crossmodal object and may make mismatches easier to detect.

The current study only considered the effect of sound delivery and movement for a task that requires crossmodal integration, so the results cannot be generalised to tasks that can be performed with vision or sound alone. Future research with a redundant audio-visual task would be useful as many multimodal displays provide redundant information (Sarter, 2006).

A possible control condition for future research might involve an existing paradigm that assesses crossmodal integration. Use of the McGurk (McGurk & MacDonald, 1976) illusion (i.e. lip-movements on the HMD and speech sounds through free-field or earpiece) might clarify whether the presently observed effect is explained by deficiencies in crossmodal integration.

Conclusion

Results show that people's movement and the method used to deliver their auditory support interact to affect people's performance of an integration task. The applied and basic literature provided very little guidance both for generating hypotheses and interpreting results. This study has addressed a practical question which has lead to the design of a novel procedure, and has created many new questions for crossmodal integration research. People move around when interpreting their multisensory environment so there is a need to better understand how crossmodal integration works when people are moving. The present findings suggest a cause for concern for people using HMDs and auditory displays in dynamic environments under some conditions. Performance with free-field sound was worse when walking than sitting. Performance with an earpiece, however, has an advantage of being insensitive to movement. Future research into different audio-visual configurations may enhance our understanding of how best to provide information to people in safety-critical environments.

References

Abernethy, B. (1988). Dual-task methodology and motor skills research: Some applications and methodological constraints. *Journal of Human Movement Studies, 14*, 101-132.

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

Alais, D., Morrone, C., & Burr, D. (2006). Separate attentional resources for vision and audition. *Proceedings of the Royal Society 273*, 1339-1345.

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology, 15*, 839 – 843.

Begault, D. R. (1993). Head-up auditory displays for traffic collision avoidance system advisories: a preliminary investigation. *Human Factors, 35*(4), 707-717.

Bertelson, P. (1998). Starting from the ventriloquist: The perception of multimodal events. *Advances in psychological science, 2*, 419–439.

Block, F. E., Jr., Yablock, D. O., & McDonald, J. S. (1995). Clinical evaluation of the 'head-up' display of anesthesia data. *International Journal of Clinical Monitoring and Computing, 12*, 21-24.

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.

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.

Calvert, G. A., Spence, C., & Stein, B. E. (2004). *The Handbook of Multisensory Processes*: Bradford Books.

Cohen, P. R., & McGee, D. R. (2004). Tangible multimodal interfaces for safety-critical applications. *Communications of the ACM, 47*(1), 41-46.

Craven, R. M., & McIndoe, A. K. (1999). Auditory monitoring--how much information do we register? *British Journal of Anaesthesia, 83*, 747-749.

De Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences, 7*(10), 460-467.

Dillon, A., & Morris, M. G. (1996). User acceptance of information technology: Theories and models. *Annual Review of Information Science and Technology, 31*, 3-32.

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.

Driver, J., & Spence, C. (1998). Cross-modal links in spatial attention. *Philosophical transactions of the royal society of London, 353*(1373), 1319-1331.

Edworthy, J., & Hellier, E. (2005). Fewer but better auditory alarms will improve patient safety. *Quality and Safety in Health Care*, *14*, 212-215.

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

Eysenck, H. J. (1980). *Manual of the Eysenck Personality Inventory*: Hodder & Stoughton.

Furnham, A., & Strbac, L. (2002). Music is as distracting as noise: the differential distraction of background music and noise on the cognitive performance of introverts and extraverts. *Ergonomics, 45*(3), 203-217.

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*(5), 473-490.

Grabowski, M., & Sanborn, S. D. (2003). Human performance and embedded intelligent technology in safety-critical systems. *International Journal of Human-Computer Studies*, *58*(6), 637-670.

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

Hodge, B., & Thompson, J. F. (1990). Noise pollution in the operating theatre. *Lancet, 335*, 891-894.

Howard, I. P., & Templeton, W. B. (1966). *Human spatial orientation*. London: Wiley.

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*(3), 967-979.

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.

Kam, P. C., Kam, A. C., & Thompson, J. F. (1994). Noise pollution in the anaesthetic and intensive care environment. *Anaesthesia, 49*(11), 982-986.

Keller, K., & Colucci, D. (1998). Perception in HMDs: What is it in Head Mounted Displays (HMDs) that really make them all so terrible? In H. J. G. R.J. Lewandowski, & L.A. Haworth (Ed.), *Helmet and Head-Mounted Displays III* (pp. 46-53): International Society for Optical Engineering.

Kramer, G. (1993). *Auditory Display: Sonification, Audification, and Auditory Interfaces*: Perseus Publishing.

Kramer, G. (1994). An introduction to auditory display. In G. Kramer (Ed.), *Auditory display: Sonification, audification, and auditory interfaces.* New York: Addison-Wesley.

Krupenia, S. S. (2007). *Monitoring and unexpected event detection with a head mounted display.* The University of Queensland, Brisbane.

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.

Lewald, J., Ehrenstein, W. H., & Guski, R. (2001). Spatio-temporal constraints for auditory-visual integration. *Behavioural Brain Research, 121*(1-2), 69-79.

Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research, 16*(3), 468-478.

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.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746-748.

Melzer, J. E., & Moffitt, K. W. (1997). *Head-mounted Displays: Designing for the User*: McGraw-Hill.

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

Miles, W. R. (1930). Ocular dominance in human adults. *Journal of General Psychology, 3*, 412–420.

Morris, R. W., & Montano, S. R. (1996). Response times to visual and auditory alarms during anaesthesia. *Anaesthesia and Intensive Care, 24*(6), 682-684.

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

Nesbitt, K. V., & Barrass, S. (2002). Evaluation of a Multimodal Sonification and Visualisation of Depth of Market Stock Data. *International Conference on Auditory Display (ICAD), Kyoto, Japan*.

Neuhoff, J. G. (2004). *Ecological Psychoacoustics*: 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*: Bradford Books.

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*(1), 17-33.

Ormerod, D. F., Ross, B., & Naluai-Cecchini, A. (2002). *Use of see-through head-worn display of patient monitoring data to enhance anesthsiologists' response to abnormal clinical events.* Paper presented at the 6th International Symposium on Wearable Computers, Seattle, Washington.

Ormerod, D. F., Ross, B., & Naluai-Cecchini, A. (2003). Use of an augmented reality display of patient safety. *Studies in Health Technology and Informatics, 94*, 248-250.

Patterson, Watts-Perotti, J., & Woods, D. D. (1999). Voice loops as coordination aids in space shuttle mission control. *Computer Supported Cooperative Work (CSCW), 8*(4), 353-371.

Patterson, Winterbottom, M. D., & Pierce, B. J. (2006). Perceptual issues in the use of head-mounted visual displays. *Human Factors, 48*(3), 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*(4), 702-715.

Sanderson, P. (2006). The multimodal world of medical monitoring displays. *Applied Ergonomics, 37*(4), 501-512.

Sanderson, P., Watson, M., Russell, W. J., Jenkins, S., Liu, D., Green, N., et al. (2007). *Advanced auditory displays and head-mounted displays: Advantages and disadvantages for monitoring by the distracted anesthesiologist*. Paper presented at the Society of Technology in Anesthesia Annual Meeting.

Sanderson, P., Watson, M. O., & Russell, W. J. (2005). Advanced Patient Monitoring Displays: Tools for Continuous Informing. *Anaesthesia and Analgesia, 101*, 161-168.

Sarter, N. B. (2006). Multimodal information presentation: Design guidance and research challenges. *International Journal of Industrial Ergonomics, 36*(5), 439-445.

Seagull, F. J., & Sanderson, P. M. The Trojan horse of the operating room: Alarms and the noise of anesthesia. *Human error in health care: A handbook of issues and indications. Mahwah, NJ: Lawrence Erlbaum Associates*.

Seagull, F. J., Xiao, Y., Mackenzie, C., & Wickens, C. (2000). Auditory alarms: From alerting to informing. *Proceedings of the Joint Meeting of the Human Factors and Ergonomics Society and the International Ergonomics Association (IEA2000/HFES2000)*, 223-226.

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport, 12*(1), 7-10.

Sparrow, W. A., Begg, R. K., & Parker, S. (2003). Aging effects on visual reaction time in a single task condition and when treadmill walking. *Motor Control, 10*(3), 201-211.

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., & Driver, J. (2004). *Crossmodal Space and Crossmodal Attention*: Oxford University Press.

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

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*: MIT Press Cambridge, MA.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The 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*.

Via, D. K., Kyle, R. R., Kaye, R. D., Shields, C. H., Dymond, M. J., Damiano, L. A., et al. (2003). *A head mounted display of anesthesia monitoring data Improves time to recognition of critical events in simulated crisis scenarios*. Paper presented at the Annual Meeting of the Society for Technology in Anesthesia.

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.

Wickens, C. D., Goh, J., Helleberg, J., Horrey, W. J., & Talleur, D. A. (2003). Attentional models of multitask pilot performance using advanced display technology. *Human Factors, 45*(3), 360-380.

Woods, D. D. (1995). The Alarm Problem and Directed Attention in Dynamic Fault Management. *Ergonomics, 38*(11), 2371-2393.

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*(3), 390-408.

Yeh, M., Wickens, C. D., & Seagull, F. J. (1999). Target Cuing in Visual Search: The Effects of Conformality and Display Location on the Allocation of Visual Attention. *Human Factors, 41*(4).

Appendix A. Mismatch Task Scenario Statistics

	Mismatches			Bounces			Ratio of Mismatches to Total Bounces (%)		
Logical Order	**Shape To Shape**	**Shape To Wall**	**Total**	**Shape To Shape**	**Shape To Wall**	**Total**	**Wall Bounce**	**Shape To Shape**	**Total**
1	1	21	22	42	255	297	7.07	0.34	7.41
2	4	29	33	69	252	321	9.03	1.25	10.28
3	9	28	37	67	260	327	8.56	2.75	11.31
4	7	20	27	56	258	314	6.37	2.23	8.6
5	8	15	23	66	262	328	4.57	2.44	7.01
6	11	24	35	64	258	322	7.45	3.42	10.87
7	5	29	34	55	260	315	9.21	1.59	10.79
8	7	16	23	69	260	329	4.86	2.13	6.99

Appendix B. Background Questionnaire

BACKGROUND QUESTIONNAIRE

Participant code: ______________	**Date: __________**	**Time:________**
Experimenter:		

Age: _________ years

Gender: Male / Female

Have you had your hearing tested? Yes / No

*If **Yes** please specify results of the test:*

____

____

If you have any hearing impediments please describe them here:

____

____

Have you had your vision tested? Yes / No

*If **Yes** please specify results of the test:*

____

____

If you have any visual impediments please describe them here:

____

____

The experimenter will help you with the following:

Dominant Eye ___________

Appendix C. Post-condition Questionnaire

Participant code: _______________	**Date: __________ Time:________**
Experimenter: ________________	**Condition: Amb-EP**

How easy was it to detect the mismatches (please circle)?

1	2	3	4	5	6	7
Very Hard ☹			Moderately Easy/Hard			Very Easy ☺

Any comments or clarifications?

How easy was it to keep count of the mismatches?

1	2	3	4	5	6	7
Very Hard ☹			Moderately Easy/Hard			Very Easy ☺

Any comments or clarifications?

How easy was it to integrate the visual information with the sounds?

1	2	3	4	5	6	7
Very Hard ☹			Moderately Easy/Hard			Very Easy ☺

Any comments or clarifications?

How easy was it to timeshare monitoring for mismatches and pushing the pushbuttons?

1	2	3	4	5	6	7
Very Hard ☹			Moderately Easy/Hard			Very Easy ☺

Any comments or clarifications?

Appendix D. Post-experiment Questionnaire

POST-EXPERIMENT QUESTIONNAIRE

Participant code: _______________	**Date: __________**	**Time:________**
Experimenter: _________________		

Please answer the following questions regarding the easiness of the different conditions. Think about how easy they were relative to each other.

	Walking, with sound coming from **speaker**	**Walking**, with sound coming from **earpiece**	**Sitting down,** with sound coming from **speaker**	**Sitting down**, with sound coming from **earpiece**
How easy was it to detect the mismatches?	1 2 3 4 5 6 7 Very Hard … Very Easy	1 2 3 4 5 6 7 Very Hard … Very Easy	1 2 3 4 5 6 7 Very Hard … Very Easy	1 2 3 4 5 6 7 Very Hard … Very Easy
How easy was it to keep count of the mismatches?	1 2 3 4 5 6 7 Very Hard … Very Easy	1 2 3 4 5 6 7 Very Hard … Very Easy	1 2 3 4 5 6 7 Very Hard … Very Easy	1 2 3 4 5 6 7 Very Hard … Very Easy
How easy was it to integrate the visual information with the sounds?	1 2 3 4 5 6 7 Very Hard … Very Easy	1 2 3 4 5 6 7 Very Hard … Very Easy	1 2 3 4 5 6 7 Very Hard … Very Easy	1 2 3 4 5 6 7 Very Hard … Very Easy
How easy was it to timeshare monitoring for violations and pushing the pushbuttons?	1 2 3 4 5 6 7 Very Hard … Very Easy	1 2 3 4 5 6 7 Very Hard … Very Easy	1 2 3 4 5 6 7 Very Hard … Very Easy	1 2 3 4 5 6 7 Very Hard … Very Easy

Any comments or clarifications?

____

____

____

____

PLEASE TURN OVER

How annoying (if at all) were the bounce sounds?

|-----------------|-----------------|------------------|-----------------|----------------|-----------------|

1	2	3	4	5	6	7
Not at all annoying ☺			Moderately annoying			Very annoying ☹

How annoying (if at all) was the button box press reminder sound?

|-----------------|-----------------|------------------|-----------------|----------------|-----------------|

1	2	3	4	5	6	7
Not at all annoying ☺			Moderately annoying			Very annoying ☹

What strategy did you use to monitor and count the violations?

____

____

____

____

What strategy did you use to timeshare monitoring for violations and pushing the push buttons?

____

____

____

____

Is there anything else you would like to tell us?

____

____

____

____

Appendix E. NASA-TLX Questionnaire

Participant code: ______________	**Date: __________** **Time:________**
Experimenter: ________________	**Condition: Sed-FF**

Mental Demand:

How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?

|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|

Low High

Physical Demand:

How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|

Low High

Temporal Demand:

How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|

Low High

Effort:

How hard did you have to work (mentally and physically) to accomplish your level of performance?

|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|

Low High

Performance:

How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|__|

Good Poor

Frustration Level:

How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Appendix F. Results of Mismatch Task

Table F1. Analysis of Variance for Mismatch Task

Source	*df*	*MS*	*F*	*partial* η^2	*p*
Movement	1	0.183	8.887	0.319	0.008**
Error	19	0.020			
Sound Delivery	1	1.017	1.122	0.056	0.303
Error	19	0.015			
Mismatch Rate	1	1.058	78.042	0.804	< 0.001***
Error	19	0.013			
Movement x Sound Delivery	1	0.148	6.449	0.253	0.020*
Error	19	0.022			
Movement x Mismatch Rate	1	0.027	2.753	0.127	0.113
Error	19	0.010			
Sound Delivery x Mismatch Rate	1	0.041	5.125	0.212	0.036*
Error	19	0.008			
Movement x Sound Delivery x Mismatch Rate	1	0.001	0.018	0.001	0.894
Error	19	0.023			

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table F2. Tukey Post-hoc Test for Mismatch Task

Condition	1	2	3	4
1. Walk-FF	-			
2. Walk-EP	0.651	-		
3. Sit-FF	0.006*	0.073	-	
4. Sit-EP	5.531	0.997	0.107	-

* $p < .05$.

Appendix G. Results of Button-press Task

Table G1. Analysis of Variance for Button-press Task

Source	*df*	*MS*	*F*	partial η^2	*p*
Movement	1	859644084	345.517	0.948	< 0.001***
Error	19	2480816			
Sound Delivery	1	11090	0.064	0.003	0.803
Error	19	173610			
Mismatch Rate	1	123750	2.013	0.096	0.172
Error	19	61464			
Button Identity	3	697992	3.116	0.141	0.033*
Error	57	224035			
Movement x Sound Delivery	1	104940	0.416	0.021	0.527
Error	19	252197			
Movement x Mismatch Rate	1	242197	3.575	0.158	0.074
Error	19	67745			
Sound Delivery x Mismatch Rate	1	207	0.003	0.000	0.961
Error	19	83566			
Movement x Button Identity	3	59431	0.352	0.018	0.788
Error	57	168693			
Sound delivery x Button Identity	3	22488	0.705	0.036	0.553
Error	57	31889			
Mismatch Rate x Button Identity	3	22601	0.539	0.028	0.657
Error	57	41924			
Movement x Sound Delivery x Mismatch Rate	1	4525	0.072	0.004	0.791
Error	19	62473			
Movement x Sound Delivery x Button Identity	3	48755	1.425	0.070	0.245
Error	57	34209			
Movement x Mismatch Rate x Button Identity	3	87837	2.009	0.096	0.123
Error	57	43716			
Sound Delivery x Mismatch Rate x Button Identity	3	18671	0.378	0.019	0.770
Error	57	49460			
Movement x Movement x Mismatch Rate x Button Identity	3	31943	0.925	0.046	0.435
Error	57	34536			

* $p < .05$. ** $p < .01$. *** $p < .001$.

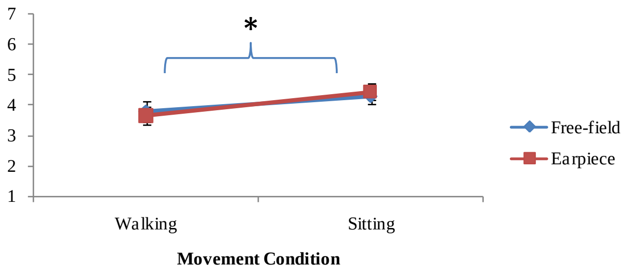
Table G2: Tukey Post-hoc test for Button-press Task

Button Identity	A	B	C	D
A	-			
B	0.597	-		
C	0.028*	0.371	-	
D	0.137	0.782	0.902	-

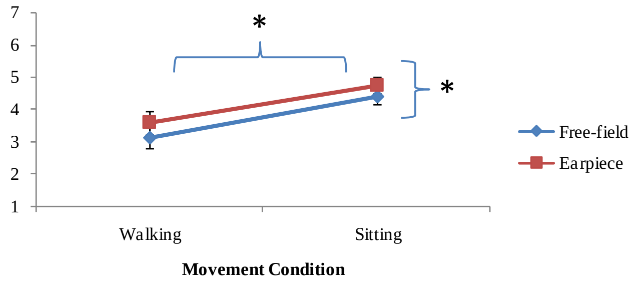
* $p < .05$.

Appendix H. Results for each question in Post-condition and Post-experiment Questionnaires

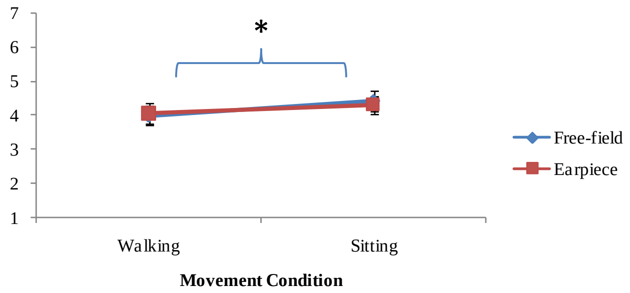
Post-condition - How easy was it to detect the mismatches?



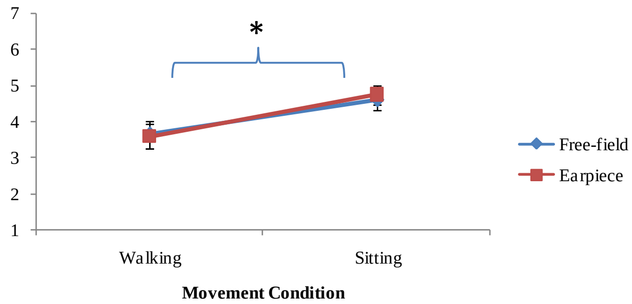
Post-experiment - How easy was it to detect the mismatches?



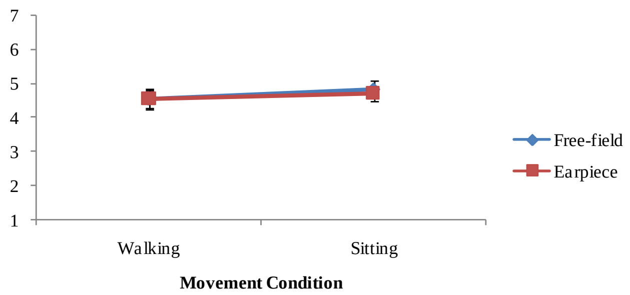
Post-condition - How easy was it to keep count of the mismatches?



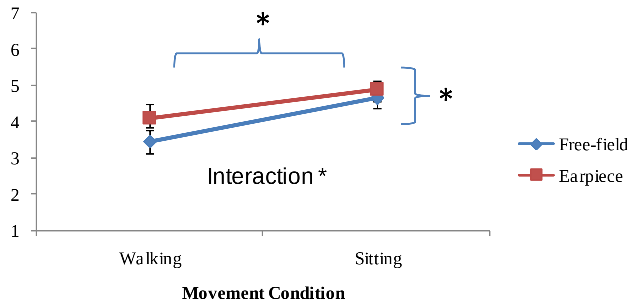
Post-experiment - How easy was it to keep count of the mismatches?



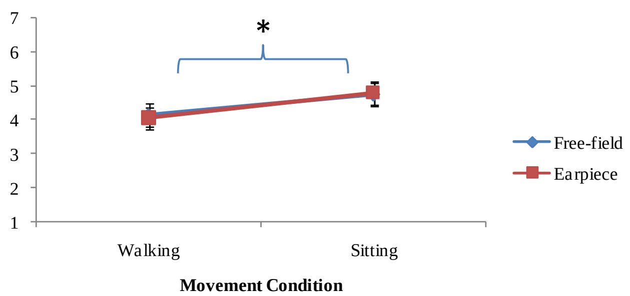
Post-condition - How easy was it to integrate the vision with the sounds?



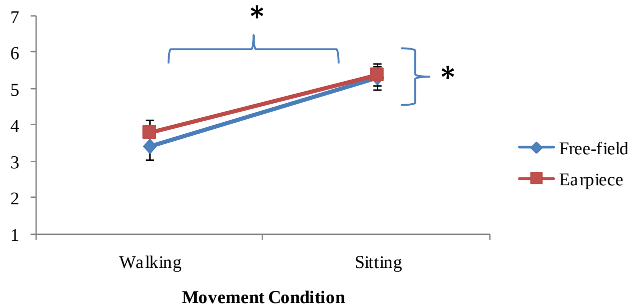
Post-experiment - How easy was it to integrate the vision with the sounds?



Post-condition - How easy was it to timeshare the two tasks?



Post-experiment - How easy was it to timeshare the two tasks?



Appendix I. Results of Questionnaire Data

Table I1. Post-condition - How easy was it to detect the mismatches?

Source	*df*	*MS*	*F*	partial η^2	*p*
Movement	1	9.592	18.853	0.512	< 0.001***
Error	18	0.509			
Sound Delivery	1	0.013	0.014	0.001	0.907
Error	18	0.930			
Movement x Sound Delivery	1	0.118	0.201	0.011	0.660
Error	18	0.591			

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table I2. Post-experiment- How easy was it to detect the mismatches?

Source	*df*	*MS*	*F*	partial η^2	*p*
Movement	1	28.800	20.886	0.524	< 0.001***
Error	19	1.379			
Sound Delivery	1	3.200	5.630	0.229	0.028*
Error	19	0.568			
Movement x Sound Delivery	1	0.050	0.322	0.017	0.577
Error	19	0.155			

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table I3. Post-condition - How easy was it to keep count of the mismatches?

Source	*df*	*MS*	*F*	partial η^2	*p*
Movement	1	3.803	4.579	0.203	0.046*
Error	18	0.830			
Sound Delivery	1	0.118	0.221	0.012	0.644
Error	18	0.535			
Movement x Sound Delivery	1	0.118	0.247	0.014	0.625
Error	18	0.480			

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table I4. Post- experiment - How easy was it to keep count of the mismatches?

Source	*df*	*MS*	*F*	partial η^2	*p*
Movement	1	14.500	9.291	0.328	0.007**
Error	19	1.555			
Sound Delivery	1	0.800	2.923	0.133	0.104
Error	19	0.274			
Movement x Sound Delivery	1	0.200	2.111	0.100	0.163
Error	19	0.095			

* p < .05. ** p < .01. *** p < .001.

Table I5. Post-condition - How easy was it to integrate the visual information with the sounds?

Source	*df*	*MS*	*F*	partial η^2	*p*
Movement	1	1.895	2.706	0.131	0.117
Error	18	0.700			
Sound Delivery	1	0.211	0.248	0.014	0.625
Error	18	0.849			
Movement x Sound Delivery	1	0.053	0.112	0.006	0.742
Error	18	0.469			

* p < .05. ** p < .01. *** p < .001.

Table I6. Post- experiment - How easy was it to integrate the visual information with the sounds?

Source	*df*	*MS*	*F*	partial η^2	*p*
Movement	1	20.000	13.333	0.412	0.002**
Error	19	1.500			
Sound Delivery	1	4.050	7.364	0.279	0.014**
Error	19	0.550			
Movement x Sound Delivery	1	0.800	5.630	0.229	0.028*
Error	19	0.142			

* p < .05. ** p < .01. *** p < .001.

Table I7. Post-condition - How easy was it to timeshare monitoring for mismatches and pushing the buttons?

Source	*df*	*MS*	*F*	partial η^2	*p*
Movement	1	9.592	10.365	0.365	0.005**
Error	18	0.925			
Sound Delivery	1	0.013	0.023	0.001	0.881
Error	18	0.569			
Movement x Sound Delivery	1	0.013	0.026	0.001	0.875
Error	18	0.513			

* p < .05. ** p < .01. *** p < .001.

Table I8. Post- experiment - How easy was it to timeshare monitoring for mismatches and pushing the buttons?

Source	*df*	*MS*	*F*	partial η^2	*p*
Movement	1	61.250	27.544	0.592	< 0.001***
Error	19	2.224			
Sound Delivery	1	1.250	4.524	0.192	0.047*
Error	19	0.276			
Movement x Sound Delivery	1	0.450	4.171	0.180	0.055
Error	19	0.108			

* p < .05. ** p < .01. *** p < .001.

Appendix J. Results of NASA-TLX

Table J1. Analysis of Variance for NASA-TLX

Source	*df*	*F*	*MS*	partial η^2	*p*
Movement	1	45.805	524.39	0.718	< 0.0001***
Error	18		11.45		
Sound Delivery	1	0.307	2.11	0.017	0.587
Error	18		6.87		
TLX-Subscale	5	16.129	634.69	0.473	< 0.0001***
Error	90		39.35		
Movement x Sound Delivery	1	1.106	6.63	0.063	0.287
Error	18		5.50		
Movement x TLX-Subscale	5	2.769	21.54	0.133	0.023*
Error	90		7.78		
Sound Delivery x TLX-Subscale	5	0.769	2.23	0.041	0.574
Error	90		2.92		
Movement x Sound Delivery x TLX-Subscale	5	0.495	1.85	0.027	0.779
Error	90		3.74		

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table J2: Tukey Post-hoc for NASA-TLX

Movement	TLX Subscale	1	2	3	4	5	6	7	8	9	10	11	12
1 Walking	Mental Demand	-	-	-	-	-	-	-	-	-	-	-	-
2 Walking	Physical Demand	0.0001***		-	-	-	-	-	-	-	-	-	-
3 Walking	Temporal Demand	0.0001***	0.0831		-	-	-	-	-	-	-	-	-
4 Walking	Effort	.3005	0.0001***	0.0002***		-	-	-	-	-	-	-	-
5 Walking	Performance	0.0016**	0.0001***	0.0378*	0.7854		-	-	-	-	-	-	-
6 Walking	Frustration	0.0001***	0.0012**	0.9685	0.0082**	0.6258		-	-	-	-	-	-
7 Sitting	Mental Demand	0.1024	0.0001***	0.0005***	1.000	0.9686	0.0378		-	-	-	-	-
8 Sitting	Physical Demand	0.0001***	0.0001***	0.0001***	0.0001***	0.0001***	0.0001***	0.0001***		-	-	-	-
9 Sitting	Temporal Demand	0.0001***	0.9999	0.3005	0.0001***	0.0001***	0.0082**	0.0001***	0.0001***		-	-	-
10 Sitting	Effort	0.0006***	0.0001***	0.0831	0.5972	1.000	0.8085	0.8872	0.0001***	0.0001***		-	-
11 Sitting	Performance	0.0001***	0.0336*	01.000	0.0003***	0.0923	0.9967	0.0018**	0.0001***	0.1517	0.1824		-
12 Sitting	Frustration	0.0001***	0.6820	0.9912	0.0001***	0.0008***	0.3483	0.0001***	0.0001***	0.9522	0.0023**	0.9421	-

* $p < .05$. ** $p < .01$. *** $p < .001$.

Appendix K. Results of EPQ

	Question	Movement	Sound	Psycho	Extrav	Neurot	Lie
Post-condition							
	Detect	Walk	FF	-.41	0.06	-0.13	0.26
	Detect	Walk	EP	-0.22	-0.21	0.12	.43*
	Detect	Sit	FF	-.56**	-0.30	-0.36	-0.55**
	Detect	Sit	EP	-.41**	-0.15	0.06	0.28
	Count	Walk	FF	-0.32	-0.18	-.43*	0.29
	Count	Walk	EP	-0.33	-.41*	-0.28	0.37
	Count	Sit	FF	-0.32	-.42*	-.45*	0.23
	Count	Sit	EP	-0.33	-0.24	-.45*	0.26
	Integrate	Walk	FF	-.41*	0.26	0.14	0.19
	Integrate	Walk	EP	-.47**	-0.08	0.24	0.27
	Integrate	Sit	FF	-.51**	-.44*	-0.36	0.24
	Integrate	Sit	EP	-.62**	-0.15	0.01	0.26
	Timeshare	Walk	FF	-.40*	0.01	-0.32	0.27
	Timeshare	Walk	EP	-.54**	-0.09	-0.09	0.37
	Timeshare	Sit	FF	-0.31	-0.19	-0.26	0.08
	Timeshare	Sit	EP	-0.21	0.05	-0.01	0.06

* Marginal

** Significant

	Question	Movement	Sound	Psycho	Extrav	Neurot	Lie
Post-experiment							
	Detect	Walk	FF	-0.25	0.18	0.19	0.04
	Detect	Walk	EP	-0.28	0.07	-0.01	0.16
	Detect	Sit	FF	-.51**	-0.19	-0.02	0.35
	Detect	Sit	EP	-.45**	-0.04	-0.17	0.38
	Count	Walk	FF	-0.19	-.42*	0.25	0.27
	Count	Walk	EP	-0.21	-0.37	0.24	0.35
	Count	Sit	FF	-0.15	-.41*	-0.02	0.30
	Count	Sit	EP	-0.03	-0.30	-0.17	0.37
	Integrate	Walk	FF	-0.13	0.21	0.11	-0.20
	Integrate	Walk	EP	-0.18	0.20	0.13	-0.07
	Integrate	Sit	FF	-0.28	-0.15	-0.14	0.23
	Integrate	Sit	EP	-.54**	-0.09	-0.02	0.36
	Timeshare	Walk	FF	-0.15	0.00	0.22	0.22
	Timeshare	Walk	EP	-0.22	0.07	0.22	0.31
	Timeshare	Sit	FF	-0.17	-0.10	-0.07	0.32
	Timeshare	Sit	EP	-0.23	-0.12	-0.12	0.37

* Marginal

** Significant

			EPQ Scale			
Movement	Sound	Mismatch Rate	Psycho	Extrav	Neurot	Lie
Walking	Free-field	Low	-0.22	-0.22	0.11	0.10
Walking	Earpiece	High	-0.19	-0.16	0.03	0.22
Walking	Free-field	Low	-0.25	0.11	-0.03	0.13
Walking	Earpiece	High	-0.35	-0.27	-0.25	0.27
Sitting	Free-field	Low	-0.36	0.16	-0.36	0.17
Sitting	Earpiece	High	-0.33	-0.05	-0.38	0.01
Sitting	Free-field	Low	-0.36	-0.10	-0.14	0.07
Sitting	Earpiece	High	-0.48**	-0.28	-0.10	0.21

* *p* < .05.

Appendix L. Information Sheet for Participants

Advanced Display Technologies in a Dynamic 360 Degree Environment

INFORMATION SHEET

In this experiment, a set of shapes will be presented on a Head Mounted Display. These shapes will be moving around the screen. Your first task is to keep a silent mental count of the number of times certain events happen on the screen based on the instructions given to you by the experimenter.

Your second task will be to move around the room pressing buttons according to the instructions given.

All data recorded in this study will be de-identified and stored according to the guidelines published by the University of Queensland's Behavioural and Social Sciences Ethical Review Committee. You are free to withdraw from the study at any time and will not be penalised in any way. If, for any reason, you do not want to continue with the experiment, simply let the experimenter know. In this event you will still be awarded full credit.

This experiment will run for approximately two hours. This study does not contain potentially offensive or embarrassing procedures or stimuli. There is a slight risk of headache from Head Mounted Display mounting band and visual fatigue but this is not beyond the risks of everyday living. A responsible UQ staff member will be available nearby during the session.

This study has been cleared in accordance with the ethical review process of the University of Queensland and is within the guidelines of the National Health and Medical Research Council. You are, of course, free to discuss your participation with project staff member Penelope Sanderson at any time (contact details below). If you would like to speak to an officer of the University who is not involved in the study you can contact the School of Psychology Ethics Review Officer directly on 3365 6394 or by email: john@psy.uq.edu.au for John McLean, or contact the University of Queensland Ethics Officer on 3365 3924, email: humanethics@research.uq.edu.au.

Experimenter: Matthew Thompson ARC Key Centre for Human Factors and Applied Cognitive Psychology McElwain Building University of Queensland, QLD 4072 Australia Tel: +61 7 3365 9510 mbthompson@psy.uq.edu.au	Supervisor: Professor Penelope Sanderson ARC Key Centre for Human Factors and Applied Cognitive Psychology McElwain Building University of Queensland, QLD 4072 Australia Tel: +61 7 3365 3988 or -6076 psanderson@humanfactors.uq.edu.au

Appendix M. Informed Consent for Participants

Advanced Display Technologies in a Dynamic 360 Degree Environment

INFORMED CONSENT

Name: ____

Address: ____

____

Email: ____

I have been provided with information about the procedure for the **Advanced Display Technologies in a Dynamic 360 Degree Environment** study and I am happy to take part. I understand I will receive two credit points in return for around an hour of participation. I understand that I can withdraw from the study at any point without prejudice or penalty of any kind. A responsible UQ staff member will be in attendance or available nearby during my session.

No notes or logs will bear any information by which I might be identified. All material will be analysed and archived under lock and key within the UQ Usability Laboratory and its confidentiality will be maintained. I also give permission for the experimental session to be video recorded for data analysis purposes.

Participant's Signature: ________________________________Date:__________

Experimenter's Signature: ________________________________Date:__________

Experimenter: Matthew Thompson ARC Key Centre for Human Factors and Applied Cognitive Psychology McElwain Building University of Queensland, QLD 4072 Australia Tel: +61 7 3365 9510 mbthompson@psy.uq.edu.au	**Supervisor:** Professor Penelope Sanderson ARC Key Centre for Human Factors and Applied Cognitive Psychology McElwain Building University of Queensland, QLD 4072 Australia Tel: +61 7 3365 3988 or -6076 psanderson@humanfactors.uq.edu.au

Appendix N. Debriefing Sheet for Participants

Advanced Display Technologies in a Dynamic 360 Degree Environment

DEBREIFING SHEET

Thankyou for participating in the **"Advanced Display Technologies in a Dynamic 360 Degree Environment"** experiment. There are two main aims of the study. The first aim is to evaluate Head Mounted Display technology in a 360 degree work environment. In a study with similar aims to this one Ormerod, Ross and Naluai-Cecchini (2003) found that Head Mounted Displays allow anaesthetists in the operating room to detect critical patient events faster than when they used traditional displays.

The study's second aim is to examine the effect of multi-modal attention (attending to both visual and auditory stimuli concurrently) on advanced display technologies. We are investigating whether auditory support for performing a task with a Head Mounted Display differs in effectiveness when the auditory support is given through an earpiece or via speakers in the room.

The results of this study have implications for the introduction of Head Mounted Displays and auditory displays into complex and safety critical systems such as operating rooms, aviation and driving.

Experimenter: Matthew Thompson ARC Key Centre for Human Factors and Applied Cognitive Psychology McElwain Building University of Queensland, QLD 4072 Australia Tel: +61 7 3365 9510 mbthompson@psy.uq.edu.au	Supervisor: Professor Penelope Sanderson ARC Key Centre for Human Factors and Applied Cognitive Psychology McElwain Building University of Queensland, QLD 4072 Australia Tel: +61 7 3365 3988 or -6076 psanderson@humanfactors.uq.edu.au