THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Evaluating head-worn displays for alarm prioritisation and situation awareness in multiple-patient contexts

Michael Thomas Pascale

MA Psychology

BA Psychology

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in 2018

School of Psychology

Abstract

Bedside patient monitoring systems in hospitals are designed to monitor a patient's physiology and to inform clinicians when a patient's vital signs exceed normal boundaries by sounding an audible alarm. The alarms, however, are widely considered to be uninformative, and consistently unreliable which can lead to inappropriate alarm management and unintended patient decompensation. Several techniques to improve the alarm systems have been tested, ranging from user-centred approaches, like improved instructions for setting thresholds, to technological improvements, like "smart" alarm systems, but no universal solution to the problem has been adopted.

Head-worn displays (HWDs) offer an alternative option: eliminating the problem by reducing reliance on bedside alarm systems entirely. Modern HWDs fit like pair of prescription glasses, but they contain a screen that could be used to display physiological information from a range of patients without the user having to physically interact with the device. Moreover, the patient information would be continuously accessible from anywhere, only requiring a rapid glance to the display, affording the user easy access to information that would otherwise be inaccessible. The HWD has the potential to reduce alarm mismanagement by creating an information rich environment, which is not possible with the current alarm systems.

The results from previous research investigating HWDs have been mixed. On the one hand, studies have shown that the additional information can improve clinicians' ability to monitor single patients for changes. On the other hand, HWDs can incur cognitive and perceptual challenges that can make the information on the display, and in the environment, less perceptible, which can worsen task performance. In addition, many current HWDs display information in the periphery, rather than over the forward field of view. What is still unknown, therefore, is how well peripheral stimuli on an HWD attract attention, and how effectively the device can be used in multiple patient contexts.

This thesis reports a series of five experiments that were conducted to explore how information on an HWD guides attention towards unexpected events. Undergraduate students, both with and without clinical experience, were tested under a variety of conditions that investigated event detection and performance while they used an HWD versus when they had to rely on standard bedside monitors in a simulated hospital microworld.

In Study 1 and Study 2, participants performed simple tasks in foveal vision and tried to detect abrupt changes to peripheral stimuli presented either on a computer monitor or on an HWD (Google Glass). The goal was to compare how abrupt changes in brightness and

apparent motion affected detection rates for peripheral stimuli presented on the two displays. The results echoed those of previous HWD studies. Specifically, detection rates for stimuli on the HWD were significantly reduced, when compared to matched stimuli on a computer monitor, especially when presented further into the periphery. However, as the stimuli on the HWD increased in distinctiveness from their context, so did the rate of detection.

In Study 3 and Study 4, participants monitored simulated patients in a computerised hospital microworld using either beside alarms alone, or bedside alarms plus a continuous stream of information presented either on a computer monitor or on an HWD (Google Glass). The goal was to explore whether the continuous stream of information improved participants' ability to prioritise clinically relevant alarms ("sick patients") over other alarms ("technological faults"). Additionally, in Study 4, I was interested in how performance on a cognitive task was affected while participants monitored their patients. Results indicated that the continuous displays significantly increased the number of relevant alarms that were managed, and reduced response times to the alarms, without affecting ongoing task performance.

In Study 5, trained participants (second and third year nursing students) monitored simulated patients in three display conditions while completing a complex patient assessment task. The goal was to test participants' ability to answer specific questions (situation awareness levels SA1 and SA2) about their patients when they were faced with either (a) alarms by themselves, (b) an HWD (Vuzix M100) and alarms, or (c) an HWD and a set of cycling auditory notifications. The results from the experiment suggested that the HWD can improve participants' situation awareness, but only when combined with the cycling display, and only in the first session. Furthermore, the HWD was associated with better performance at the patient assessment task.

Taken together, the findings suggest that HWDs have the potential to bring useful information to a clinician, allowing them to focus on the task at hand with less interruption. The additional information that the HWD provides may be useful for disambiguating alarms and maintaining continuous awareness of the status of multiple patients, which is not possible with the current design of bedside alarms. Appropriately designed HWDs may improve hospital ward environments by allowing clinicians to recognize negative trends sooner and to prioritise the management of sick patients over other tasks.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, financial support and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my higher degree by research candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis and have sought permission from co-authors for any jointly authored works included in the thesis.

Publications during candidature

Conference abstracts

Pascale, M., Sanderson, P., Liu, D., Mohamed, I., Loeb, R. (2015). Event detection using a simulated head-worn display. *The Proceedings of the 19th Triennial Congress of the International Ergonomics Association.* Paper presented at the 19th Triennial Congress of the IEA: IEA2015, Melbourne (p. 1-3).

Pascale, M., Sanderson, P., Liu, D., Mohamed, I., Stigter, N., Loeb, R. (2015). Peripheral detection for abrupt onset stimuli presented via head-worn display. *Proceedings of the 59th Annual Meeting of the Human Factors and Ergonomics Society.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting: HFES 2015, Los Angeles (p. 1326-1330). Sage Publications.

Pascale, M., Sanderson, P., Liu, D., Mohamed, I., Brecknell, B., Loeb, R. (2016). Continuous information displays for multiple patient monitoring. *Proceedings of the 60th Annual Meeting of the Human Factors and Ergonomics Society.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting: HFES2015, Washington DC (p. 1556-1556). Sage Publications.

Manuscripts included in this thesis

Pascale, M., Sanderson, P., Liu, D., Mohamed, I., Stigter, N., Loeb, R. (under review). Detection of visual stimuli on monocular peripheral head-worn displays. *Applied Ergonomics. Submitted 13 December 2017.*

Contributor	Statement of contribution
Pascale, Michael (candidate)	Conception and design (55%) Analysis and interpretation (70%) Drafting and production (65%)
Sanderson, Penelope	Conception and design (25%) Analysis and interpretation (10%) Drafting and production (20%)
Liu, David	Conception and design (5%) Analysis and interpretation (20%) Drafting and production (5%)
Mohamed, Ismail	Conception and design (5%) Drafting and production (5%)
Stigter, Nicola	Conception and design (5%)
Loeb, Robert	Conception and design (5%) Drafting and production (5%)

Pascale, M., Sanderson, P., Liu, D., Mohamed, I., Brecknell, B., Loeb, R. (2017). Impact of head-worn displays on strategic alarm management. *Human Factors. Revise and resubmit, 7 August 2017.*

Contributor	Statement of contribution
Pascale, Michael (candidate)	Conception and design (55%) Analysis and interpretation (65%) Drafting and production (60%)
Sanderson, Penelope	Conception and design (25%) Analysis and interpretation (10%) Drafting and production (20%)
Liu, David	Conception and design (5%) Analysis and interpretation (25%) Drafting and production (5%)
Mohamed, Ismail	Conception and design (5%) Drafting and production (5%)
Brecknell, Birgit	Conception and design (5%)
Loeb, Robert	Conception and design (5%) Drafting and production (10%)

Contributions of others to the thesis

Professor Penelope Sanderson was Mr. Pascale's principal PhD advisor. Professor Sanderson made significant contributions to the development of the theoretical framework, experimental design, data interpretation, the writing of the manuscripts submitted for publication, and the writing of this thesis.

Dr. David Liu was the associate advisor for this work. Dr Liu contributed to experimental design, development of statistical frameworks, data analysis, and data interpretation.

Dr. Robert G. Loeb was a research collaborator throughout this body of work. Dr Loeb contributed his clinical expertise to theoretical development, experimental design, and preparation of manuscripts.

Statement of parts of the thesis submitted for the award of another degree

None.

Research Involving Human or Animal Subjects

The research complied with the Australian National Statement on Ethical Conduct in Human Research and was approved by the Human Research Ethics Committee of the School of Psychology at The University of Queensland (approval 14-PSYCH-PHD-59-JS). Informed consent was obtained from each participant.

Acknowledgements

Above all, I would like to thank my PhD advisor Professor Penelope Sanderson. Thank you for sharing your passion for human factors research, and for being patient with me as I maneuverer through various stages of this thesis. Most of all, thank you for devoting so much time to my education and for helping me find potential within myself.

A big thank you to my associate advisor Dr. David Liu for guiding me in experimental design, and giving me the tools to find meaning in data. Also, to Dr. Robert Loeb, for contributing your clinical expertise to the design of my studies.

Big hugs to my CERG-sisters: Tara, Mia, Chiara, Kelly, and Estrella. Each of you has contributed to the progress of my PhD in a different way, whether it be through encouragement, frequent reality checks, mutual frustration, or simply good old-fashioned home baked brownies. Thank you to Ismail Mohamed and Birgit Brecknell, for tending to my never-ending stream of feature requests and bug reports while you developed the software I needed to do my research.

To the staff at the Princess Alexandra Hospital; thank you for letting me follow you around for days on end asking questions. Also to the staff within The School of Nursing and Midwifery at UQ; thank you for giving me the opportunity to pitch my research to your students, and for encouraging them to volunteer.

Last but not least, thank you to my brother Andrew. I have always looked to you for direction and guidance, and this PhD and Aussie experience has been no different. And, to my sisters, Leilani and Lelia; thank you for always supporting me on my adventures, even though I know you both wish I was closer to home.

Financial support

The research presented in this thesis was funded by ARC Discovery Project Grant DP140101822 (Sanderson, Loeb, & Liu) and a University of Queensland International Scholarship, with further support from the School of Psychology at UQ.

Keywords

cognitive engineering, human factors, head-worn displays, perception and performance, situation awareness, patient monitoring

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 170112, Sensory Processes, Perception and Performance, 70%
ANZSRC code: 179999, Psychology and Cognitive Sciences not elsewhere classified, 30%

Fields of Research (FoR) Classification

FoR code: 1799, Other Psychology and Cognitive Sciences, 80%
FoR code: 1701, Psychology, 20%

Dedications

To my father Connie Pascale, for being a constant source of inspiration and for always supporting me unconditionally.

Table of Contents

List of Figures

List of Tables

Lists of Abbreviations used in thesis

HWD Head worn display

HR Heart rate

BP Blood pressure

SpO2 Oxygen saturation

SAGAT Situation Awareness Global Assessment Technique

Q-ADDS Queensland Adult Deterioration Detection System

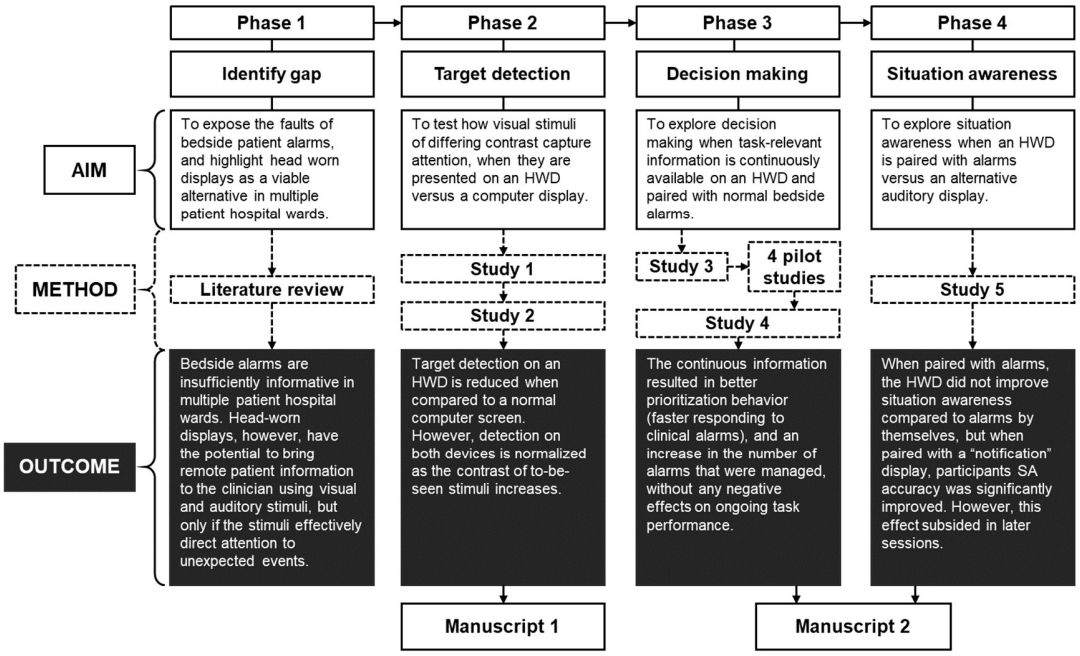

Figure 1. Overview of program of research

The purpose of this thesis was to explore whether head-worn displays could be used to help clinicians monitor multiple patients in hospitals. The goal of the experiments reported was to show that modern head worn computers could be used by participants, both with and without clinical experience, to maintain their situation awareness of the status of multiple simulated patients while they performed other tasks. A further goal was to reduce participants' reliance on bedside monitoring alarm systems by introducing more informative auditory displays.

In Phase 1, I describe the current issues with bedside monitoring systems in many hospitals. This section covers the problems associated with alarms in hospitals, the behavioural issues that stem from the problems, the body of research discussing potential solutions to the problems, and the fact that few studies explicitly consider multiple patients as a factor in alarm management. I then discuss why alarms fail to guide attention, followed by a brief description of current display techniques that do. Next, I introduce head worn displays (HWDs) as well as their pros and cons with respects to behaviour, attention, and perception, followed by the potential for HWDs to be used as monitoring devices. Finally, I outline a series of studies designed to test the attention guiding capabilities of HWDs.

In Phase 2, I report Studies 1 and 2, in which participants attempted to detect peripherally displayed stimuli while performing tasks in foveal vision. The goal was to explore how detection for visual stimuli presented on a peripheral HWD compares to stimuli presented peripherally on a standard computer display. Results suggested that HWDs impose costs to attention capture, particularly when the to-be-seen stimuli are further in the periphery and low in distinctiveness.

In Phase 3, I report Studies 3 and 4, in which participants had access to continuous streams of patient information that they could use to prioritise sick patients over technical faults in a simulated hospital microworld. The goal was to explore whether the continuous information contributed to the decision-making process, resulting in faster, and more responses to sick patients, even when the participant had to perform multiple tasks. Results from this phase provided evidence that the continuous information on the HWD could be used to disambiguate alarms, and to generate rapid responses, more often.

In Phase 4, I report Study 5, in which I tested nursing students' ability to maintain continual awareness for changes to multiple patients' status in a simulated hospital microworld, while they also performed a complex simulated patient assessment. The goal was to compare situation awareness in three display conditions: (a) alarms only, (b) alarms and an HWD, and (c) earcon notifications and an HWD. The data showed that the earcon notifications and HWD resulted in greater accuracy on situation awareness questions, but only under certain conditions in our simulated environment.

Finally, I discuss the overall results of the program of research, including theoretical and practical implications with respects to HWDs in multiple patient hospital wards.

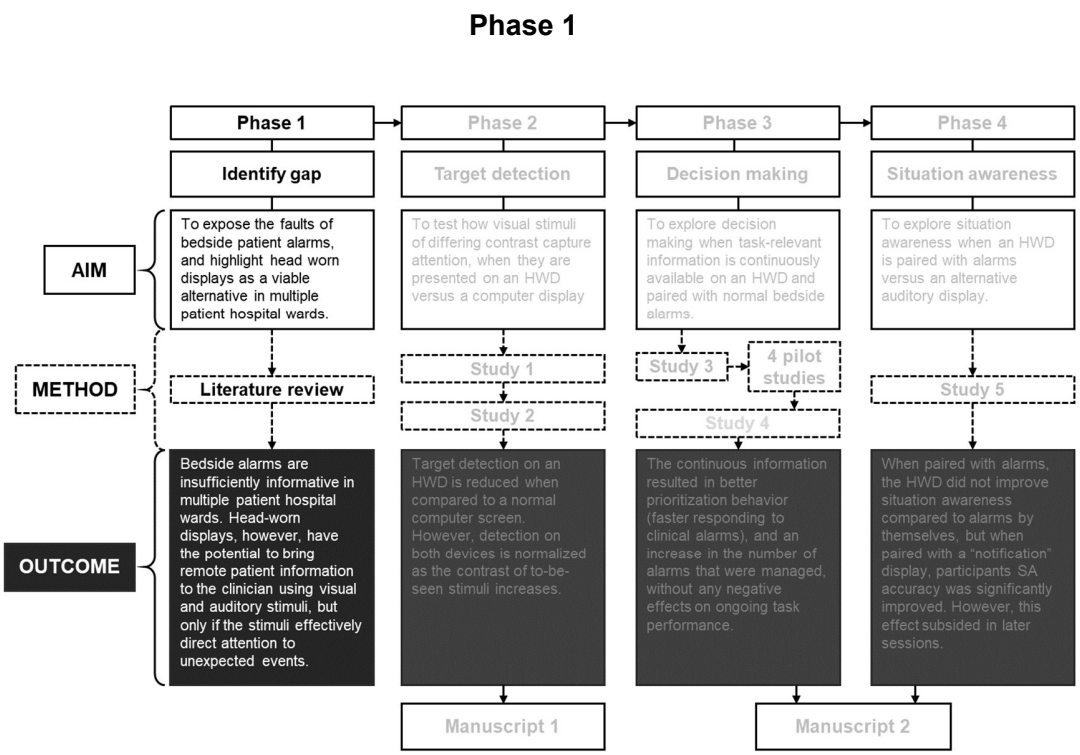

Figure 2. Phase 1 of the research program.

The purpose of this phase is to introduce the alarm problem, how it affects clinicians who are responsible for multiple patients, and how modern display technologies can be used to mitigate alarm mismanagement. I first describe how alarms that are designed to alert and inform clinicians of potentially life-threatening changes fail to do so, and then I outline current solutions. Second, I examine the alarm problem from a directed attention perspective, and describe how different displays can direct attention more reliably. Third, I introduce head-worn displays (HWDs) with an explanation of their potential to bypass the alarm problem in multiple patient monitoring contexts by providing easily accessible streams of information. Finally, I present an overview of a series of studies directed at testing the attention-directing capabilities of HWDs.

Alarms and patient monitoring systems

In this section, I discuss the current state of bedside monitoring systems and patient alarms in hospitals. First, I describe the concerns associated with alarms from a multiple patient monitoring perspective, which is a factor not often considered in hospital-alarm literature. I then discuss several attempts that have been made to solve alarm-related issues, followed by a novel solution in the form of advanced, head-worn displays.

Alarms are designed to attract an observer's attention to an event, usually one that requires immediate attention, but in many environments, alarms can be uninformative and unreliable. Essentially, the presence of alarms is not enough to initiate a response to an adverse event because alarms do not always accurately represent an adverse event. Moreover, during time periods where the alarm is supposed to be the most useful, when a clinician's task loading is highest for example, the alarm is actually a distraction and more difficult to interpret (Woods, 1995). Woods (1995) defined this as the "alarm problem" which has been documented in a variety of environments including nuclear power (Mumaw, Roth, Vicente, & Burns, 2000), air traffic control (Wickens et al., 2009), and healthcare (Paine et al., 2016). In each of these environments, uninformative alarms can lead to errors.

In hospitals, for example, the bedside monitoring systems are designed to inform the clinical staff when a patient's vital signs are moving away from acceptable levels, but the alarms are often triggered for other reasons (Table 1). The purpose of the alarms is to interrupt the clinical staff, alerting them to an important event by capturing their attention using sound, and sometimes light, at which point the clinician can decide how to respond. From an interruptions perspective, if the monitoring system were perfectly accurate, and alarms were triggered only when a patient needed attention, the interruption would benefit the clinician's work flow (Sasangohar, Donmez, Trbovich, & Easty, 2012). In this example, the clinician's decision process would be simple: if there is an alarm, a patient is in distress and needs immediate attention. Unfortunately, the alarms often occur needlessly, and do not require an intervention. Thus, when considering the variety of tasks and other interruptions that clinicians are faced with (Li, Magrabi, & Coiera, 2011; Rivera-Rodriguez & Karsh, 2010), dropping a task or tasks at hand to respond to unreliable alarms is not a top priority and might have a negative impact on patient safety, rather than a positive one.

Table 1

Taxonomy of alarm types from Ruskin and Hueske-Kraus (2015)

Alarm type	Description
Clinical alarms	Indicate that a patient's physiological state requires attention
-True clinical alarms	Patient physiology has changed and requires attention
-False clinical alarms	Patient physiology has not changed but the alarm has been triggered
-Nuisance alarms	Patient physiology has changed but does not require attention
Technical alarms	Indicate that the monitoring equipment requires attention
-True technical alarms	No signal from device
-Avoidable technical alarms	Equipment has not been appropriately set up
-False technical alarms	Signal not reliable (artefact)

Part of the reason that false and nuisance alarms are such a concern is that they occur very frequently in many hospital wards, and much more often than true alarms, creating a “cry wolf” situation. Studies have found that 80 to 99 percent of bedside alarms in hospitals are nonactionable (Cvach, 2012; Ruskin & Hueske-Kraus, 2015), meaning that a patient’s bedside alarm was triggered, but the patient did not actually require immediate attention. Furthermore, hospital wards house multiple patients in close proximity, resulting in simultaneous, nonactionable alarms coming from multiple sources. The cause of single instances of alarms is difficult enough to identify using the alarm sound, but when alarms come from patients in wards where each nurse is responsible for multiple patients (sometimes as high as 1:8, or 1:12 overnight), the alarms become even more difficult to localize and to interpret. Interestingly, however, multiple patient monitoring is not often explicitly mentioned as a factor in the alarm literature.

As the number of patients in a ward increases, thereby increasing the rate of nonactionable alarms, so do negative attitudes towards alarms, their frequency, and their consistent unreliability (Sowan, Tarriela, Gomez, Reed, & Rapp, 2015). The high rate of nuisance alarms and subsequent negative attitudes towards them can lead to inappropriate responses (Cvach, 2012; Sowan & Reed, 2017), or in some cases, a complete failure to respond to the alarms (Xiao, Seagull, Nieves-Khouw, Barczak, & Perkins, 2004). If an inappropriate response to an alarm happens when a patient is actually in distress, the risk of further complications increases, which can result in the worsening of a patient’s condition, and in some cases, death (Xiao et al., 2004). Sadly, when an avoidable patient-related accident occurs, the blame is often placed on the individual staff who were responsible, without considering these systemic factors (Davidson, Agan, Chakedis, & Skrobik, 2015; Woods, 1995), which can contribute to high nurse-burnout rates and job dissatisfaction (Ruskin & Hueske-Kraus, 2015; Solet & Barach, 2012; Taenzer, Pyke, & McGrath, 2011).

It is widely known that the high rate of patient alarms is related to the thresholds being used for the physiological monitoring systems (Paine et al., 2016). In most cases, the thresholds for each vital sign are set by the manufacturer and left untouched when attached to the patient. The default alarm thresholds are set so that no potentially threating changes to a patient’s condition are missed, but this leads to numerous false positive alarms. For example, a patient’s heart rate slows to 58 beats per minute and triggers an alarm (alarm threshold = 60), but that patient’s normal resting heart rate is 63 beats per minute. Essentially, the device is not set according to the patient’s specific physiology, which results in alarms that have no clinical relevance. Potential solutions include improving clinician knowledge of the monitoring devices, establishing routine practices, and setting individualized thresholds, which have been effective in some cases (Block

III & Block Jr, 2015; Solet & Barach, 2012), but these changes do not sufficiently increase the safety of the alarm systems (Sowan, Gomez, Tarriela, Reed, & Paper, 2016).

A technological alternative to threshold manipulations is making improvements to the artificial intelligence of the alarm systems, but that is similarly unreliable. With 'smarter' alarm systems, algorithms are used to trigger the alarm if multiple physiological symptoms are present rather than the alarm being triggered when a single vital sign exceeds a predetermined threshold (Imhoff, Kuhls, Gather, & Fried, 2009). As with threshold settings, however, the algorithms may still miss a true patient deterioration event, and it also puts some of the decision making in the 'hands' of the technology rather than those of the clinician. Recent, and future, advances in artificial intelligence may result in better smart alarm systems, but for the moment these systems are not trustworthy enough to employ safely.

From a different perspective, the risks from current alarm systems could be reduced by increasing the number of staff members. Patient safety is increased with nurse to patient ratios closer to 1:1 and when shifts are shorter (Bonafide et al., 2017). Additionally, faster recognition of patient deteriorations is seen when hospitals employ "monitor watchers" whose task is to continuously attend to physiological patient signals (Funk, Parkosewich, Johnson, & Stukshis, 1997; Watkins, Whisman, & Booker, 2016). Increases in staff, however, require additional funds, which may not be available in many hospitals.

A further area of research investigates the effectiveness of mobile devices as aids for patient monitoring, but such devices either do not bring enough information to the clinician or they have other drawbacks. Cvach, Frank, Doyle, and Stevens (2014) found that the use of pagers was associated with better nurse awareness of high-priority patient alarms, but their research simultaneously investigated an algorithm using alarm delays, which by itself has been found to be effective for reducing alarm frequencies (Gorges, Markewitz, & Westenskow, 2009). Moreover, in most cases, simple paging systems would be redundant with the alarm system, simply bringing the alarm to the clinician's body without increasing the amount of information being conveyed. A more informative method is to use tablet computers that can make patient information more accessible (Baig, GholamHosseini, & Lindén, 2015), but handheld devices can impose infection hazards and require physical manipulation to be useful, which are undesirable in sterile and hands-on hospital environments.

A head worn display (HWD), however, may be the mobile solution that bypasses the alarm problem entirely. Given the capabilities of modern display technology, and the mobile and hands-off nature of the device, the HWD may combine the benefits of continuous monitoring and of mobility, while still leaving decisions in the hands of clinicians instead of algorithms, and without forcing clinicians to rely on uninformative bedside alarms. Specifically, the HWD may help to

reduce ambiguity when multiple alarms occur simultaneously, encourage better prioritization of tasks, and improve the clinician's awareness of changes to a patient's condition. Altogether, HWDs may have the potential to improve patient outcomes and reduce clinician workload by reducing reliance on uninformative bedside alarms in wards with high nurse to patient ratios.

Directed attention with advanced displays

There are many factors that contribute to the alarm problem in hospitals which is why no universal solution has been adopted (Wilken et al., 2017). Regardless of the cause, the result is a growing distrust in the alarm systems, which can result in undesirable alarm management, and unwanted patient outcomes. If clinicians were not consistently engaged in other tasks, their attention would not be divided, and they would be more likely to recognise and respond to alarms as they were triggered, regardless of the cause. Unfortunately, clinicians *are* consistently engaged in a multitude of tasks that require focused attention, and that often should not be interrupted. It is in these situations that the current bedside alarms are least useful, because even though the alarms may attract the clinician's attention, the contextual information required to generate an appropriate response to the alarm, if deemed necessary, is not available.

As currently implemented, auditory patient alarms support preattentive orienting towards important events, but due to the high rate of false alarms such orienting often happens needlessly. Preattentive orienting, or preattentive reference, refers to the cognitive processes that precede focal attention, but that guide attention towards stimuli in the environment that match an expectation or goal, or that are potentially interesting (Woods, 1995). Auditory alarms can orient attention via preattentive reference, even when a person's attention is directed elsewhere and while they are under high cognitive or perceptual load (Horrey & Wickens, 2004; Wickens, Dixon, & Seppelt, 2005) which makes auditory alarms particularly useful in hospital environments where there can be high levels of risk if patient events go unnoticed. Unfortunately, this also means that most alarms capture attention[1] away from the task at hand, even when no response is necessary, as is the case with the vast amounts of false and nuisance alarms.

[1] Attention capture can be described as a bottom-up, stimulus-driven process (Yantis & Jonides, 1984), but also as a top-down, goal-driven process (Folk, Remington, & Johnston, 1992). The latter has been described as contingent attention capture, whereby stimuli with certain properties that meet a set of expectations or goals capture attention. In other words, with contingent attention capture preattentive selection is mediated by the expectations of the observer. Throughout this thesis, and in particular in Phase 2, all references to attention capture refer to the contingent capture hypothesis from Folk et al. (1992). See Theeuwes, Olivers, and Belopolsky (2010) for a discussion of the different modes of attention capture.

After capturing the clinician's attention, the alarms can convey, also unreliably, the urgency of an event using a few different sound parameters (Edworthy, Loxley, & Dennis, 1991). Again, this has the potential to be meaningful, but in auditory environment dominated by false and nuisance alarms, clinicians must guess whether or not the urgency being conveyed is accurate, and then act accordingly. What the current alarm standards cannot convey is the exact context of the alarm, such as why the alarm was triggered, where the alarm is sounding, or whether a vital sign is too high or too low and by how much. Thus, the alarm may capture attention, but does little more to provide support for task prioritization, requiring the clinician to seek further information before making a treatment decision. Furthermore, without any context, the clinician may mistake a true alarm for a nuisance alarm, since they are indistinguishable to the ear. Most of the solutions to this problem aim to fix the alarms themselves, but there are several alternatives to bedside alarms that have been found to be more effective for guiding attention to unexpected events.

A body of research investigating ways to increase the amount of information being conveyed by the auditory displays in hospitals has had some success. Edworthy et al. (2017) tested the alarm sounds recommended in the current IEC 60601-1-8 alarm standard against a variety of different auditory displays including earcons (sounds that represent the hazard), word rhythms (sounds that mimic the syllables of words), as well as different prioritization and localization techniques. Their data showed that the current alarm standards required longer to recognize and locate than the newer displays. Further studies have increased the amount of information being conveyed by using advanced auditory (sonification) displays that map onto, and change according to, a single patient's current condition. These studies found that using more descriptive pitch, harmonics, and tremolo improve participants' ability to recognize abnormal trends when compared to less descriptive displays (Hinckfuss, Sanderson, Loeb, Liley, & Liu, 2016; Loeb & Fitch, 2002; Paterson, Sanderson, Paterson, & Loeb, 2017; Watson & Sanderson, 2004). Additionally, Li et al. (2017) tested compressed-speech based earcons, or spearcons, and found that participants could rapidly learn the meanings of the spearcons, to identify them more accurately, and with more confidence, than normal earcons. However, that study was conducted with Cantonese speaking participants.

There is also evidence for advanced auditory displays to be effective in multiple patient contexts. In particular, Hickling, Brecknell, Loeb, and Sanderson (2017) found that participants using cycling earcon displays identified the amount, and location, of abnormal vital signs in simulated patients. Also, as Li et al. (2017) found, Bell (2017) found that English speaking participants using spearcons were able to describe the status of multiple patients, but their accuracy decreased as the number of patients increased. Together, these results suggest that

there may be better options than alarms for informing clinicians about the specific status of a patient or multiple patients. Most importantly, using an advanced auditory display would provide the clinician access to specific patient information that would otherwise be unavailable without physically moving to that patient's bedside or a centralised monitor.

One caveat with the majority of sonification literature is that the auditory devices have been localized to patient's bedsides, nursing stations, or remote speakers. Recent advances in high-powered lightweight computers, however, have encouraged a substantial increase in the potential for these displays to follow the clinician (Sanderson, 2006). There is also evidence that combining visual displays with redundant auditory displays can improve visual detection of vital sign deviations (Seagull, Wickens, & Loeb, 2001). This is where an HWD might be most useful, by bringing advanced multimodal displays directly to the clinician, regardless of the clinician's location. With such a display, the clinician's attention would be guided more effectively towards potentially adverse events, increasing awareness for changes to a patients' conditions, and improving their ability to respond rapidly in the event of a life-threatening patient deterioration.

Head worn displays

In the previous section, I described the limitations of alarms that contribute to their frequent ineffectiveness. I then discussed a body of research that tested alternative auditory displays that show promise. In this section, I will introduce HWDs and discuss how they can be used to supplement auditory displays, by bringing even more detailed information directly to clinicians.

Practical applications of HWDs have traditionally been limited by their form factor, but with recent mass production they are now small enough to be used in clinical environments such as hospital wards. Older HWDs consisted of multiple components - reflective screen, display module, computer and communications module, battery, and house - which were packaged in large assemblies and very heavy when worn on the head for long periods of time. However, as laptops, phones, and tablets are increasingly miniaturized, the form factor of each component of the HWDs has reduced in size. This trend was led by substantial research and design investment, as well as mass production runs, by Google for their Glass™ HWD. Other manufacturers, such as Vuzix, have also produced HWDs with small and light form factors. The recent improvements in form factor have resulted in renewed interest in testing the clinical applications of HWDs.

With physical aspects of the device no longer a concern, recent HWD research has been able to focus on people's performance, perception, and attention while using HWDs, with mixed results for clinical contexts. On the positive side, HWDs have improved single patient monitoring in anaesthesiology by allowing clinicians to maintain visual orientation towards the patient (Liu,

Jenkins, & Sanderson, 2009; Ormerod, Ross, & Naluai-Cecchini, 2002), and assisting in the detection of a patient deterioration that may have otherwise gone unnoticed (Vorraber et al., 2014). HWDs can simplify access to historical patient information during patient exams (Monroy, Shemonski, Shelton, Nolan, & Boppart, 2014), improve communication (Muensterer, Lacher, Zoeller, Bronstein, & Kübler, 2014) and be used as a recording device for teaching purposes (Knight, Gajendragadkar, & Bokhari, 2015). Moreover, clinicians wearing HWDs have been considered them to be comfortable, easy to read, and not distracting (Drake-Brockman, Datta, & Ungern-Sternberg, 2016). On the negative side, however, clinical information on HWDs can be missed even though it is displayed directly in the forward field of view (Dixon et al., 2013; Liu, Jenkins, Sanderson, et al., 2009) and can lead to underperformance and overconfidence while using the device (Liu, Jenkins, Sanderson, et al., 2009).

Outside the healthcare domain, HWDs have been shown to improve information acquisition during some physical tasks, but not without affecting performance on other tasks. HWDs can improve participants' ability to read and send messages while driving, but not without costs to driving performance (He, Ellis, Choi, & Wang, 2015; Sawyer, Finomore, Calvo, & Hancock, 2014; Tippey, Sivaraj, Ardoin, Roady, & Ferris, 2014). During a rock climbing task, some visual stimuli presented on an HWD were missed, but some of the information loss was eliminated when visual stimuli were paired with simultaneous auditory cues (Woodham, Billinghurst, & Helton, 2016). Moreover, during manual assembly tasks, HWDs have been associated with reduced performance when compared to pen and paper instructions (Büttner, Funk, Sand, & Röcker, 2016).

A further concern with HWDs relates directly to limitations of the visual system, but the extent of the concern depends on the form of the device and the type of stimuli being displayed. The most commonly cited issue is binocular rivalry, or the unintentional switching of the priority of visual input streams from one eye to the other. Binocular rivalry is most prominent for monocular HWD displays where augmented information is being viewed by only one eye (Patterson, Winterbottom, Pierce, & Fox, 2007; Winterbottom, Patterson, Pierce, Covas, & Rogers, 2006; Winterbottom, Patterson, Pierce, & Taylor, 2006). The concern is that information can be missed if the information is being presented to one eye while the other eye is being prioritized (Patterson et al., 2007). Similarly, there are additional problems associated with eye-dominance (LaFleur, Draper, & Ruff, 2001), and focal depth (Winterbottom, Patterson, Pierce, Covas, & Winner, 2007) both of which can affect perception for items being presented via HWD. For unexpected and transient visual stimuli, these limitations may negatively affect the attention capturing qualities of the display. However, if the user is deliberately checking the display for changes, and therefore intentionally switching priority to the HWD, the information may be less likely to go unnoticed.

Furthermore, if to-be-seen visual stimuli are paired with redundant auditory stimuli, then the likelihood of missing information is further reduced, even if the information transient and abrupt (Seagull et al., 2001; Woodham et al., 2016).

HWDs and attention for patient monitoring

In this section, I will describe how HWDs could be effectively deployed in a multiple patient monitoring context by discussing how advanced HWDs can guide attention. As outlined above, alarms that occur frequently, and that come from multiple sources simultaneously, do not provide clinicians with the information they need to respond to unexpected patient deteriorations. In fact, because the alarms are so uninformative, the clinicians sometimes respond in ways that do not contribute to the safety of their patients. However, HWDs and advanced auditory displays can be designed to increase the accessibility of information and guide the clinician to recognize patient events faster and with more context. Having access to real-time information might be the difference between responding immediately to a clinically relevant alarm versus responding seven minutes later, as illustrated in Bonafide et al. (2017).

Most importantly, an HWD offers access to remote information that may otherwise be inaccessible unless the clinician is at the patient's bedside or at a nurses' station. For a display to be most effective in a multiple patient monitoring context, it should reduce response time to adverse patient events and it should increase situation awareness by bringing easily accessed information to the clinician when they need it. The ideal set of stimuli would guide the user's attention to the display via both exogenous and endogenous pathways, and allow the user to rapidly acquire information during short, infrequent gazes at the display.

Exogenous attention capture—or attention captured involuntarily by stimuli (Folk, Remington, & Wright, 1994)—would be most useful for alerting a clinician to a critical patient event. For example, the clinician might immediately perceive that an alarm has been triggered, but the information on the display suggests that it is just a technical alarm rather than a true change in the patient's status. Thus, the clinician could delay an intervention until after the current task is completed. Alternatively, a patient's respiration rate might drop dangerously low, triggering an alarm; but the stimuli on the HWD could attract the clinician's attention towards the display so that they can perceive the change and respond immediately. The alert, in this case, must sufficiently capture the clinician's attention from the task at hand so that the information is not missed. However, increases in task loading, particularly if the task is perceptually demanding, can limit the ability of stimuli to capture attention (Lavie, 2005; Lavie, Beck, & Konstantinou, 2014). Previous studies suggest that abruptly displayed stimuli with high colour contrast (Jonides, 1981; Nikolic, Orr, & Sarter, 2004; Nikolic & Sarter, 2001; Yantis & Jonides, 1984, 1990), and

stimuli that appear to be in motion (Folk et al., 1994), can reliably break through tasking loading to capture attention, but these stimuli may also be distracting if they attract attention too often, or unnecessarily. So, first, it would be useful to know how different stimuli displayed on a monocular HWD capture attention, and how that compares to attention capture using computer displays.

Endogenous attention – or attention directed by goals, expectations, and values (Folk et al., 1994) is derived naturally over time, depending on the salience of the information, the importance of the information to the viewer or listener, and the likelihood of an unwanted or an unexpected event (Sheridan, 1970; Wickens, Goh, Helleberg, Horrey, & Talleur, 2003). The HWD, in this case, would provide continuous real-time information to clinicians, so they can maintain awareness of the status of the patients under their care. However, as patients' conditions worsen over time, the clinician's attention towards the display will change accordingly, probably resulting in more frequent sampling of the display (Senders, 1964; Sheridan, 1970). For example, the clinician might notice that a patient's respiration is slightly lower than normal. A few minutes later, the clinician might recheck the display to see if the patient has worsened or improved. The clinician might recognize that the negative trend has continued, therefore reacting before the patient reaches a critical condition. Essentially, the clinician can deliberately use the display to update their mental model of their patients' status, which would not be possible using standard patient alarms, or without physically moving to that patient's bedside or a central monitoring system.

Program of research

For the reasons given above, HWDs may help clinicians monitor the status of multiple patients, and reduce their reliance on alarms. However, there has been no direct evidence that supports this conjecture. The goal of the next phase was to test this speculation in a series of studies.

Most current HMDs offer a continuously-present display in the periphery of vision, where it might be used to capture attention via exogenous cueing. What is still unknown is whether stimuli displayed on an HWD capture attention more often or less often than matched stimuli displayed on a standard computer screen. Thus, the first two experiments (Phase 2, Studies 1 and 2) in the series were designed to test participants' ability to detect simple changes to stimuli in peripheral vision, when they were presented via monocular HWD or via a computer display. Many studies have been conducted regarding exogenous, peripheral attention capture, but it is not clear how the results of these studies carry over to monocular HWDs. Moreover, there have been no studies directly comparing peripheral attention capture on HWDs to peripheral attention capture on computer displays. I expected that stimuli presented on the HWD would be detected less

often than stimuli presented on the computer display, due to the perceptual deficits associated with HWD use, such as binocular rivalry. Specifically, I expected that the binocular view of the ongoing perceptual task would be prioritised over the monocular view of the stimuli on the HWD, resulting in less sensitivity to target changes with the HWD.

The second set of experiments (Phase 3, Studies 3 and 4) was designed to test participants' ability to use an HWD to make patient treatment decisions when faced with multiple alarms. The goal was to determine whether using an HWD would improve participants' ability to discriminate different kinds of alarms and increase the likelihood that they prioritise sick patients. I predicted that the HWD would encourage better treatment decisions, operationalized by faster response times to critical alarms, without affecting performance on an ongoing task.

The final experiment (Phase 5, Study 5) was designed to test whether there was a situation awareness advantage when trained participants (nursing students) used HWDs alongside alarms compared to alarms by themselves. Moreover, I tested whether pairing an advanced auditory display with an HWD further increases situation awareness over the HWD and alarms. In this experiment, the participants had to monitor six simulated patients for physiological changes while completing a complex patient assessment task. I predicted that participants would have greater awareness in both HWD conditions over the alarms only condition. I also predicted that the advanced auditory display and the HWD would result in greater situation awareness than the HWD and alarms.

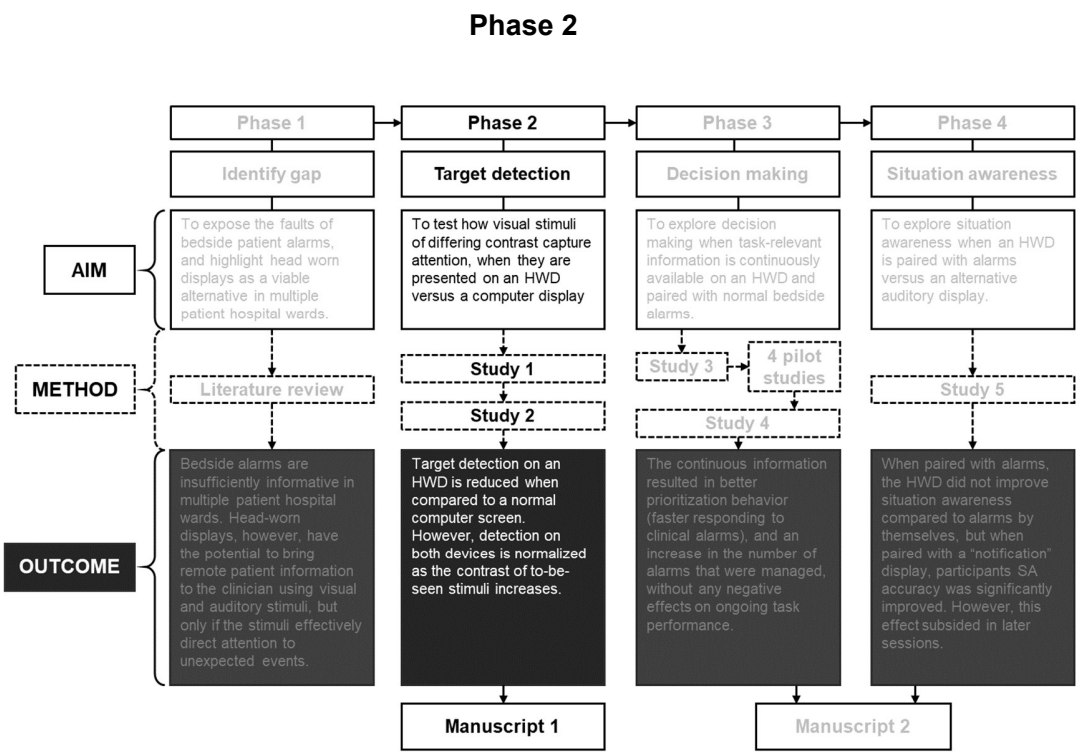

Figure 3. Phase 2 of the research program.

The purpose of Phase 2 was to test our hypotheses regarding HWDs and peripheral target detection. In two studies, participants performed an ongoing task in foveal vision and attempted to detect simple changes to stimuli in peripheral vision. The first study was completed entirely on a standard computer display, and was designed to compare detection rates for stimuli in near versus far peripheral vision, and while participants performed either a cognitive or perceptual task in foveal vision. The data from the first study were then used to design Study 2, which compared detection rates for near versus far peripheral stimuli on a monocular HWD to matched stimuli on the standard computer display, while participants completed only the perceptual task in foveal vision. These studies were written up as short conference papers (Pascale, Sanderson, Liu, Mohamed, & Loeb, 2015; Pascale, Sanderson, Liu, Mohamed, Stigter, et al., 2015), but a full writeup of Study 2 can be found as Manuscript 1 in Appendix A.

Study 1: Effects of cognitive load versus perceptual load on target detection using a computer display

The purpose of the first study in the series was to test participants' ability to detect a series of abrupt changes in brightness ("whiteness") and apparent motion (stimulus orientation, "tilt") presented in a fixed location in peripheral vision (Figure 4). Participants attempted to detect these

changes while performing either a cognitive or perceptual task in foveal vision. All stimuli in Study 1 were presented to both eyes via a computer display. I chose to first test these factors on the computer display to see if they are sensitive in the way I expected, and to see whether the accuracy rates were appropriate for me to measure whether the HWD (in Study 2) made things either better or worse.

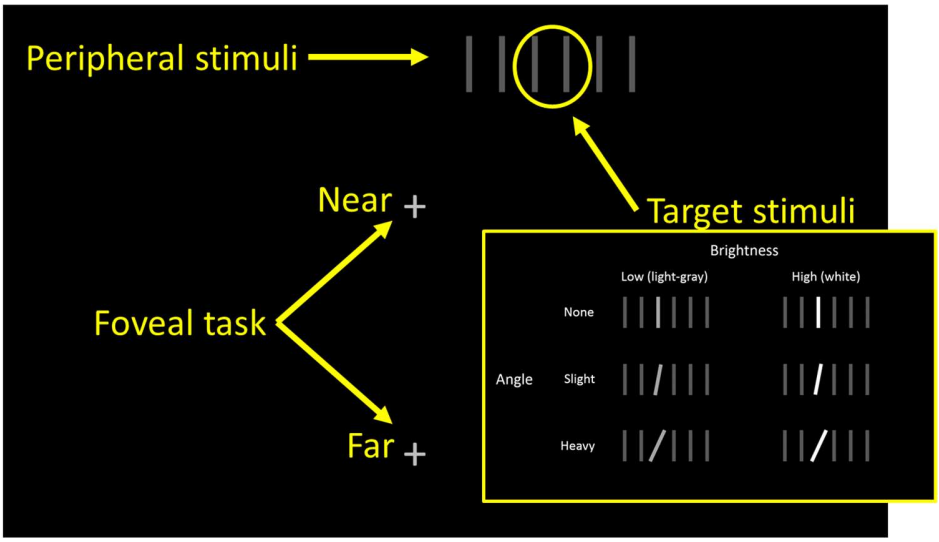

Figure 4. The layout of the screen in Study 1, showing the location of the peripheral stimuli, the two possible locations of the foveal task, and the six possible target stimuli.

Stimulus detection on computer displays is affected by the distinctiveness of the stimuli (Folk et al., 1994; Yantis & Jonides, 1990), the distance of the stimuli from foveal vision (Nikolic et al., 2004), and the requirements of the ongoing task (Lavie, 2010). What is still unknown, however, is whether peripheral stimuli on an HWD are capable of capturing attention, and how detection rates compare to those seen using a computer display. For example, are changes to target stimuli more, or less, likely to be missed when presented on an HWD compared to identical stimuli presented on a computer display? Before testing this directly, however, I needed to establish a paradigm that allowed me to measure differences in detection rates for target changes to peripheral stimuli.

Participants were seated at a fixed distance from a 21-inch computer display with their head stabilized using a chinrest. The peripheral stimuli were presented in a fixed location above and to the right of the ongoing task, simulating the location of the peripheral HWD to be used in later experiments. The peripheral stimuli consisted of six dark grey, vertical bars presented side by side. These bars were always visible, but from time to time, one of the two central bars would rapidly change states from the default dark grey colour and default vertical orientation to one of the six target stimuli. I chose 200 ms for the duration of target presentation because it was fast

enough that participants should not have been able to make a saccade to the peripheral display when they noticed the target change. Considering that the targets occurred intermittently and at seemingly random intervals to the participant, all target detection should have occurred in peripheral vision. The target stimuli were generated by making changes in brightness alone, or by making changes in both brightness and tilt.

For each block of trials, the ongoing task was either a cognitive task (arithmetic) or a perceptual task (Landolt C comparisons) presented at one of two locations on the computer display. The two different locations generated the near and far eccentricity conditions of the peripheral stimuli (see Figure 4). All participants saw four blocks of each task and each eccentricity (counterbalanced). The participant's goal was to focus on and respond to the ongoing task using the left hand to make key presses, and to monitor for changes to the stimuli displayed in peripheral vision, using the right hand to make key presses.

In Study 1, I expected the following: (1) participants would detect more target stimuli while they performed the cognitive task than the perceptual task, as found in Lavie (2005), (2) target stimuli presented at the far eccentricity would be detected less often than target stimuli presented at the near eccentricity, as found in Nikolic et al. (2004) and Wolfe, O'Neill, and Bennett (1998), and (3) the stimuli with the most distinctive features from the dark grey vertical default (e.g. white and heavily tilted) would be detected most often, as found in Yantis and Jonides (1984).

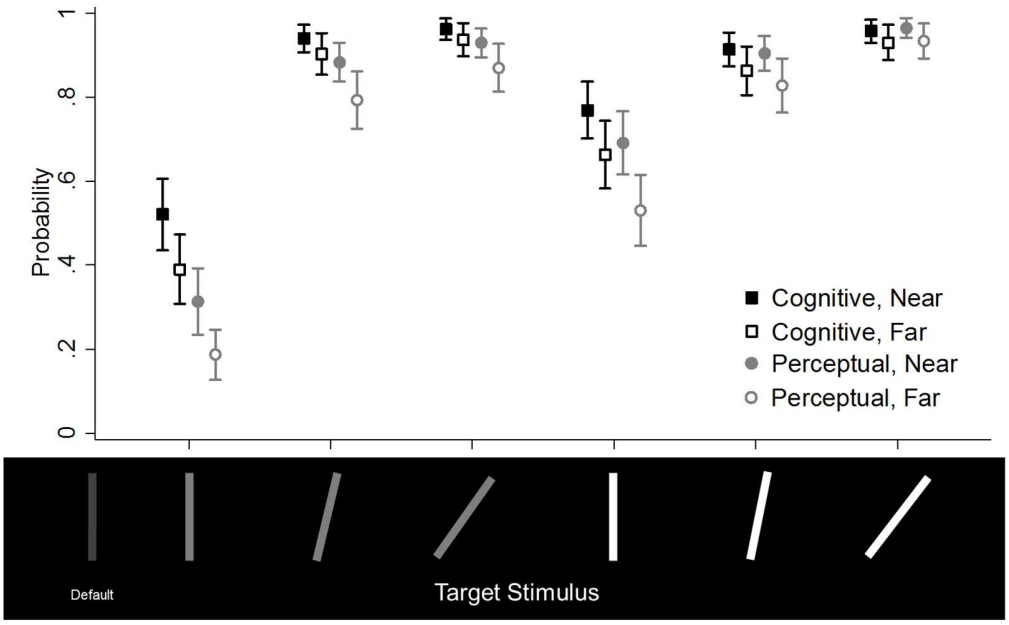

Figure 5. Predicted probability and 95% confidence intervals for detecting each target stimulus in Study 1, while participants performed either the cognitive or perceptual tasks at either the near or far eccentricity.

A mixed effects logistic regression was conducted to measure detection differences between the cognitive and perceptual tasks, the near and far eccentricities, and for each of the six target stimuli. Participants (N=15) detected peripheral stimuli less often while performing the perceptual task than when performing the cognitive task. Moreover, participants detected the stimuli less often when they were presented at the far eccentricity compared to the near eccentricity. The suppression of detection rates due to task loading and eccentricity, however, lessened as stimuli increased in distinctiveness. The predicted probabilities are shown in Figure 5. These findings were not novel, but the study established a foundation with which to test how detection rates of stimuli in a peripheral, but fixed location on a computer display compare to detection rates on HWD.

Study 2: Target detection: computer display versus head-worn display

The purpose of Study 2 (Manuscript 1, Appendix A) was to test whether participants' detection rates for abrupt changes to peripheral stimuli on a computer display, as in Study 1, were greater than their detection rates for matched changes on a real HWD. HWDs can impose perceptual constraints on the user (Winterbottom, Patterson, Pierce, Covas, et al., 2006; Winterbottom, Patterson, Pierce, Gaska, & Hadley, 2015) but it is still unknown whether simple changes to visual stimuli on HWDs can capture attention to the same degree that they would on computer displays. Thus, for Study 2, I modified the paradigm established in Study 1, by eliminating the cognitive task, which resulted in the highest rate of target detection, and eliminating the stimuli that were heavily tilted. An additional modification was a target change that consisted of just tilting the dark grey stimulus without increasing the colour contrast.

Before conducting Study 2, I needed to calibrate the stimuli so that their appearance was matched as closely as possible across the two displays (Manuscript 1 in Appendix A). This was necessary because the HWD in question (Google Glass, Google Inc., Mountain View, CA, USA) has a see-through display that cannot produce a pure black background that matches the black background of a computer display. A separate set of participants performed the calibration study, which consisted of a psychophysical "method of adjustment" task. This task required participants to wear the HWD, to maintain the position of the display over the black background of the computer display, and to adjust the colours and contrasts of each stimulus item so that they were matched across displays. Specifically, the participants used button presses to adjust the colour (redness, greenness, blueness) of the background of peripheral stimuli on the computer display to match the background colour of the Google Glass display. The participants then made adjustments to the colour and opacity of the default and target stimuli on Google Glass to match the stimuli on the simulated HWD. This task was completed in two blocks for each participant,

first with the numerical colour and opacity values visible (starting well above the expected range of values), and the second time with the numerical values hidden (starting well below the expected range of values, or vice versa), so that participants could not just use the number values to match their output on the first block.

With the stimuli matched across display devices, I could now test detection performance across the two display modes with minimal potential confounds related to stimulus presentation. Below, I refer to the stimuli on the computer display as the "simulated HWD" to account for the matched size, colour, and peripheral location as Google Glass (the "real HWD"). I expected the following three outcomes in Study 2.

(1) Target detection rates for target stimuli on the real HWD would be lower than the detection rates for target stimuli on the simulated HWD due to the visual limitations when stimuli are presented to only one eye;
(2) Target detection rates for stimuli presented at the far eccentricity would be lower than detection rates for target stimuli presented at the near eccentricity because of the reduction in retinal receptors in the periphery; and
(3) Target detection rates would increase, on both displays, as the stimuli increased in distinctiveness from the dark grey, vertical default, e.g. more brightness, or tilt.

Study 2 was conducted following the same procedure as Study 1, except participants now also had a forehead rest, to further stabilise their heads so that the real HWD was consistently viewed over the black background of the ongoing task display.

A mixed effects logistic regression showed that participants' (N=61) detection rates for stimuli presented on the real HWD were reduced compared to their detection rates on the simulated HWD, but the differences in detection rates were nearly eliminated for targets at the near eccentricity that were distinctly different from the default vertical and dark grey stimulus (Figure 6). The results supported our expectation that less distinct stimuli presented on the HWD would capture attention less often than stimuli presented on the computer display, particularly when the stimuli were presented at the far eccentricity. These findings suggest, however, that visual stimuli with high distinctiveness could be designed to alert an HWD user to important information on the display. Still, detection rates are likely to be further affected by increased levels of task loading, as well as greater eccentricities of the visual stimuli in the periphery.

The results of Phase 2 established that the HWD can make it difficult for participants to notice target stimuli that are not distinctive, especially when those stimuli are presented further in the periphery. This is a factor that any designer of HWD displays must consider to ensure that time sensitive information is not missed. For the remaining studies in this thesis, I will no longer be manipulating visual distinctiveness, but the to-be-seen stimuli have been designed following the

guidelines above, using high colour contrast. Moreover, in Phase 4 I also test an auditory means of capturing attention exogenously to guide visual attention to the display.

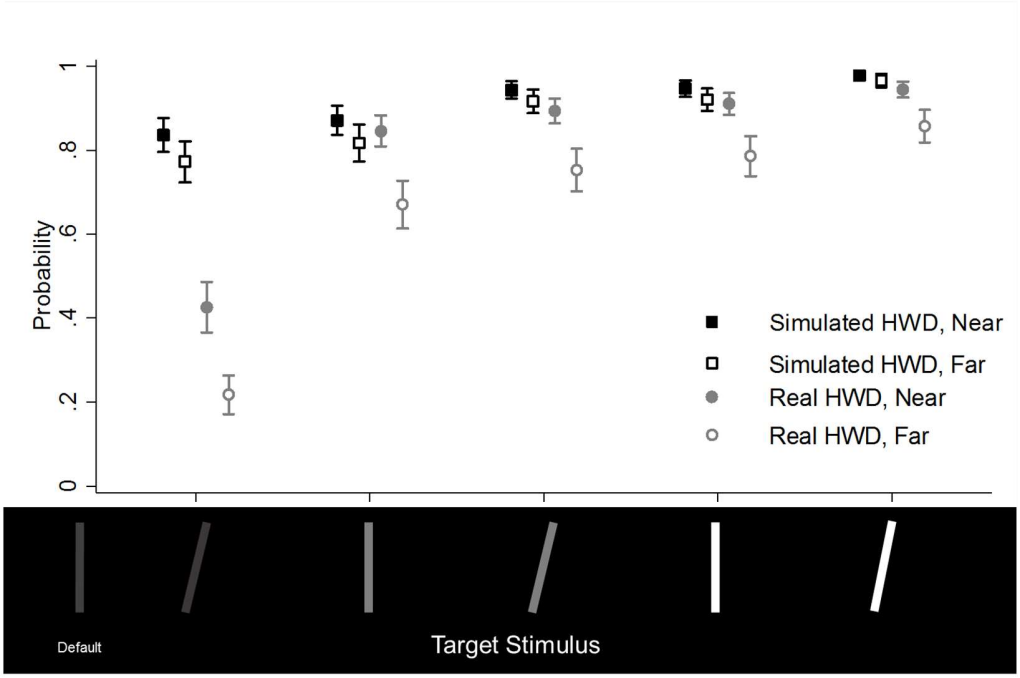

Figure 6. Predicted probability and 95% confidence intervals for detecting each target stimulus in Study 2, presented on either the simulated HWD or Google Glass (real HWD), at either the near or far eccentricity.

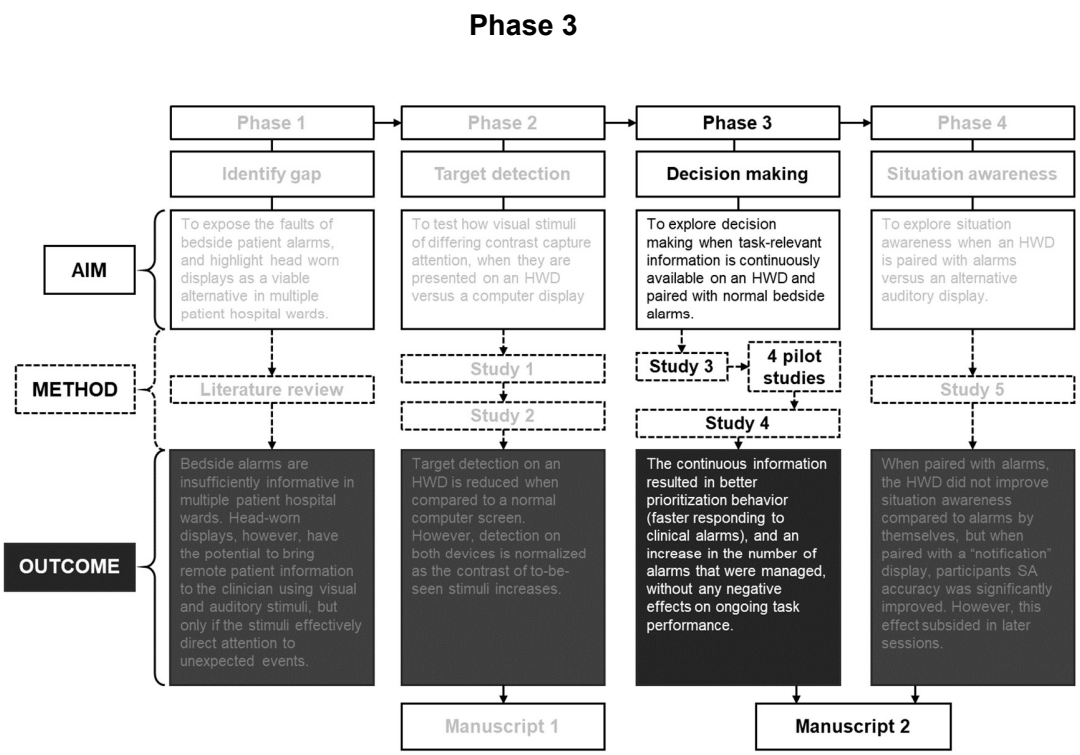

Figure 7. Phase 3 of the research program.

The purpose of Phase 3 was to take a few steps closer to a clinical situation and test our hypotheses regarding the potential effect of HWDs on participants' ability to manage alarms. In two studies, participants monitored simulated patients in a computerised microworld of a hospital ward (Figure 8) and responded to "bedside" patient alarms. Study 3 investigated participants' ability to prioritise the management of clinically relevant alarms over nonclinical alarms, by relying on either in-room, bedside monitors (top of Figure 8) or by relying on bedside monitors and a continuous screen-based display of all the patients' vital signs (bottom of Figure 8), and moving to the appropriate patient room. Study 4 was conducted similarly, but the continuous display was shown on an HWD, instead of on the same screen as the microworld, and the patients had to complete an ongoing "dosage calculation" task. Phase 3 tested whether the continuous displays improved participants' decision making compared to when they had to rely on the alarms without any further information, as is typically the case in many hospital wards.

Study 3: Multiple patient monitoring in a simulated hospital microworld

In Phase 2, Study 2 showed that visual stimuli on an HWD are more susceptible to being missed, but it is unlikely that rapidly displayed, and transient, visual stimuli would solely be used in an environment where attention needed to be directed to events immediately. What is

commonly seen (or rather, heard) in high risk environments is auditory alarms designed to attract the attention of the people nearby, or the people connected with the alarm system wirelessly. In this context, the HWD may act as a supplement to the alarm system by displaying task relevant information from remote sources, rather than acting as the primary method for capturing attention towards an adverse event. Thus, after attracting attention to the display, the HWD would assist the user in deciding how to respond to any alarms that were triggered.

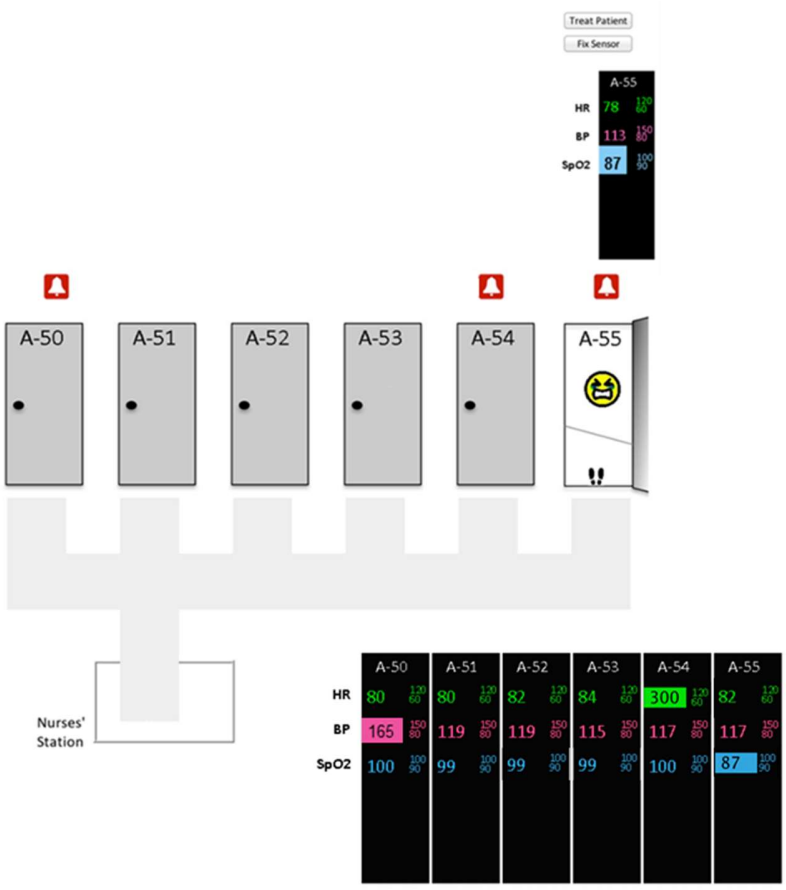

Figure 8. View of the computerised microworld in the alarms-monitor condition, showing six patient rooms represented by the grey doors and the continuous display of all patients' vital signs (bottom right). In the alarms only condition, the continuous display was not visible to participants. The feet indicate the participant's location in the simulated "ward". In this case the patient is in room A-55. The red icons above the door indicate that a bedside alarm is on.

In hospital wards where the average response time to alarms is often more than seven minutes (Bonafide et al., 2017), and when watching a central monitor continuously is not possible (Funk et al., 1997; Watkins et al., 2016), having continuous access to patient data might

encourage clinicians to make better decisions when a clinically relevant, actionable alarm sounds, and thus encourage faster response times to patient deteriorations. For example, a nurse checking on the status of one patient might hear an alarm suggesting that another patient is deteriorating. After checking the HWD, the nurse might recognize that the pattern of data does appear to be in line with cardiac arrest. The nurse could then immediately excuse herself from her current task (if safe to do so) to attend to the deteriorating patient or call for help.

Study 3 was designed to test how participants with continuously accessible information displays made patient prioritisation decisions, compared to participants who had to use alarms alone. The task and stimuli for Study 3 were designed to approximate some of the issues faced by nursing staff in multiple patient hospital wards. Specifically, I used a computerized hospital microworld that could be manipulated in a variety of ways. I could control the number of patients, their vital signs, distances between rooms, walking pace for the simulated nurse, alarm criticality (whether each alarm was clinically relevant versus irrelevant), and the alarm thresholds. Most importantly, I could control whether information from all patients was continuously available to the participant (alarms only versus alarms and monitor), as in Figure 8.

The alarm criticality was driven by the numerical values of the vital signs, which were explained to participants during the instruction and practice phases. Specifically, participants were instructed that alarms triggered by a vital sign with a value of zero or 300 suggested that the alarm (a sensor alarm) had been triggered because of a technological malfunction, rather than a “physiological” deterioration, which did not require an immediate response. However, an alarm triggered by a vital sign with a plausible value (a clinical alarm) that was too high or too low was described as clinically relevant, which did require an immediate response. The plausible ranges for each vital sign were posted at the task computers.

The participants’ goal was to respond to, and prioritize, clinical alarms to reduce the amount of time any patient was left in an adverse state. Participants navigated the microworld using the arrow keys on a standard keyboard, and managed alarms by entering the representation of a patient’s room and clicking either a “Treat patient” button for clinically relevant alarms, or a “Fix sensor” button for the irrelevant, technological alarms (top right of Figure 8). Each participant completed three blocks, each lasting 10 minutes. There were 22 clinical alarms and 36 sensor alarms in each block with an ISI of 10.4 seconds.

The purpose of Study 3 was to test whether participants in the alarms-monitor condition could respond to a larger proportion of the clinically relevant alarms, and respond faster, than the participants in the alarms only condition. I chose first to establish the benefits of having patient information continuously available on a conventional screen before investigating any benefits of having it continuously available on an HWD. If I did not first establish that participants benefit

from continuous information, and if performance with the HWD were to show no benefit, I would not be sure whether it was because the HWD was ineffective or because the continuous information itself conferred no benefit. I hypothesized the following.

(1) Participants in the alarms-monitor condition would respond to clinically relevant patient alarms faster than participants in the alarms only condition (treatment latency).

(2) Participants in the alarms-monitor condition would respond to a greater number of clinically relevant alarms, than participants in the alarms only condition (untreated alarms).

A linear regression revealed that participants (N=24) using the continuous information display in the alarms-monitor condition could indeed respond to patient alarms faster (left image of Figure 9) than participants using alarms only. Additionally, a Poisson regression revealed that participants in the alarms-monitor condition left fewer clinical alarms untreated (right image of Figure 9) than participants in the alarms only condition. The results of Study 3 therefore suggest that continuous information displays can be used to prioritize high priority patients in the event of multiple alarms, corroborating the results of Funk et al. (1997) and Watkins et al. (2016).

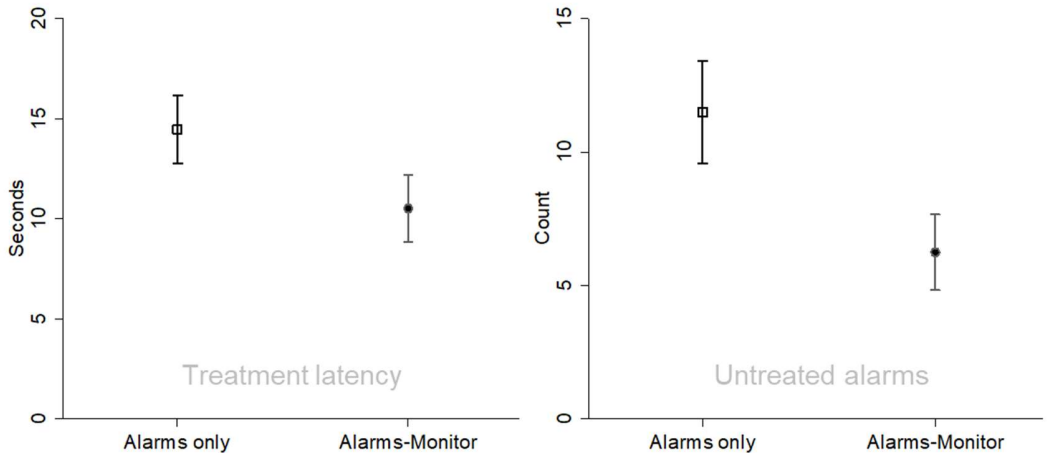

Figure 9. Predictive models with 95% confidence intervals produced from the regression analyses in Study 3.

Study 4: Effects of head-worn display on multiple source monitoring and ongoing task performance

In Study 3, I established that participants could use a continuous display to disambiguate alarms, and to prioritise relevant alarms over irrelevant alarms, and that such a display gave them an advantage over a display that showed only the location of alarms. For Study 3, the continuous display was shown on the same computer screen as the microworld.

In Study 4, I was interested in whether the benefits of having access to the continuous information persisted when the information was presented on an HWD and while the participants had to complete an ongoing task in a separate location from the microworld (Figure 10). In the alarms only condition, participants had to stand up and go to the microworld to investigate alarms. In the alarms plus HWD condition, the participants could look on the display to see whether it was likely to be a clinical or technical alarm, and plan accordingly.

Given the results of Study 2, which showed a reduction in the likelihood that stimulus changes are noticed when they are on an HWD, it was uncertain whether the HWD would automatically lead to patient prioritisation improvements, as in Study 3, or if limitations from the device would make the information less usable, especially when an additional task was introduced. Study 4 was included as Experiment 1 in Manuscript 2 (see Appendix B).

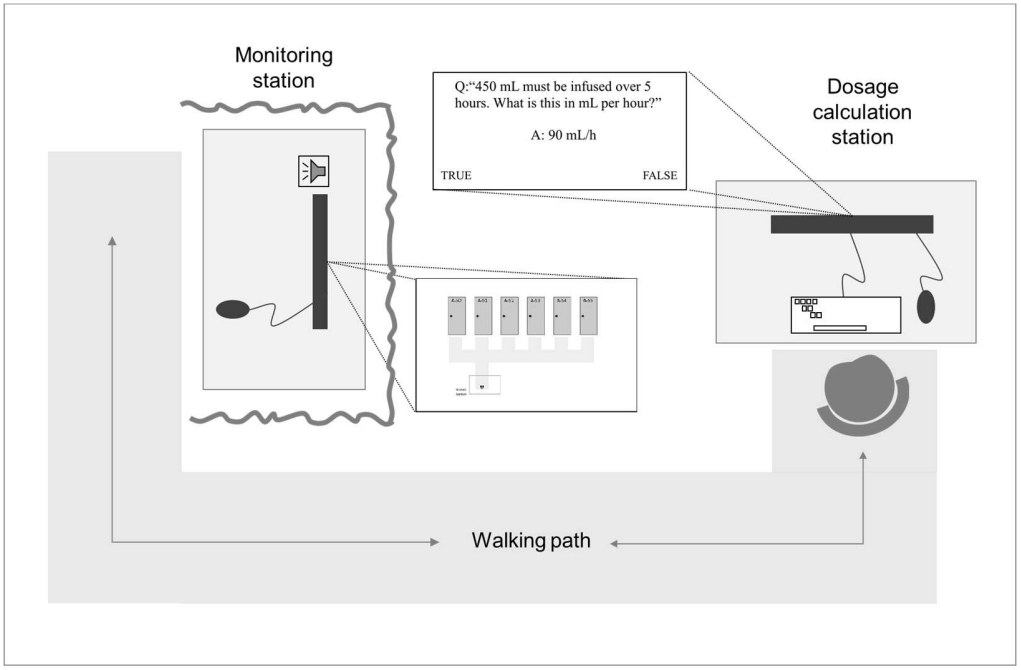

Figure 10. Diagram of the layout for Study 4. The monitoring station contained the participant's view of the microworld. The ongoing task was completed at the dosage calculation station.

The purpose of Study 4 was to test whether the HWD influenced participants' ability to prioritize patient alarms in the hospital microworld. The study was conducted in much the same way as Study 3, but the continuous information display was now presented on the HWD (HWD-alarms condition) rather than on the computer screen showing the hospital microworld. Additionally, I included an ongoing task that was in a separate location in the test room from the hospital microworld. I included the ongoing task so that participants' attention was not solely engaged on monitoring the display or moving between patients. Similarly, including the ongoing

task allowed me to test how monitoring with the HWD might affect secondary task performance, or vice versa.

The ongoing task, a dosage calculation task, required participants to read word problems, do simple multiplication and division in their head, and to make a response regarding whether the given answer to the problem was correct (True) or incorrect (False) (see Figure 10 for an example). The participants' goal was to complete as many dosage calculation questions as possible, while monitoring for and treating clinically relevant patient alarms.

Table 2

Summary of pilot studies conducted prior to Study 4. All are between-subjects designs.

			Number of alarms (probability)		Treatment latency	Dosage time	Dosage accuracy
Pilot	N	Alarm frequency	Clinical	Sensor	*p*	*p*	*p*
1	10	-	12 (40%)	18 (60%)	0.69	0.4	0.2
2	12	-	8 (25%)	24 (75%)	0.26	.003*	0.16
3	10	-	13 (29%)	32 (71%)	.006*	0.2	0.08
4	22	Low	9 (30%)	21 (70%)	< .001*	.016*	0.62
		High	13 (29%)	32 (71%)			

Note. All *p*-values are from ANOVAS investigating the effect of display condition.

Prior to conducting Study 4, I completed a series of pilot experiments. All pilot studies were fully between-subjects designs comparing performance in the alarms only condition to performance in the HWD-alarms condition. The participants' goal in each pilot was to monitor patients for clinical alarms, to treat them promptly, and to complete as many dosage calculations as they could. Summaries of the four pilot studies can be found in Table 2.

The first pilot (N=10, with n=5 per condition) was designed to test whether the new dual-task paradigm, as well as the complex set of instructions regarding task and alarm prioritisation, could elicit the task prioritisation behaviours I anticipated. Data from Pilot 1 suggested that the participants were not fully understanding the prioritization, because they stopped performing the ongoing task when both types of alarms occurred, rather than just when the clinically relevant alarms occurred.

The second pilot (N=12, with n=6 per condition) consisted of a similar study but with modified instructions to more explicitly describe the alarm prioritization behaviour. Specifically, participants were now instructed that they could "safely ignore" the irrelevant sensor alarms, to focus on responding to clinically relevant patient alarms and completing dosage calculations. I also modified the number of alarms, the ratio of clinical alarms to sensor alarms (see Table 2), and

included a few opportunities for participants to practice alarm prioritisation. Data from Pilot 2 suggested that participants were now appropriately prioritizing clinical alarms, but that the effect was not very strong. I conjectured that the HWD may not have improved alarm prioritisation because there were too few alarms, so that participants in the alarms only condition needed to move to the microworld screen only infrequently, and still manage to treat all the clinical alarms.

The third pilot (N=10, with n=5 per condition) was conducted to test how participants' performance differed when alarms occurred more frequently. This pilot study was conducted using the same design as Pilot 2, but the number of alarms for each 10-minute block was increased from 32 to 45. Furthermore, the schedule of alarms was modified so that there was always an active alarm, representing the worst-case scenario for a clinician. I hypothesized that the increase in alarms would make the information on the HWD more useful for prioritising clinical alarms. The data from Pilot 3 confirmed this, showing that participants with the HWD were leaving fewer alarms untreated, and treating clinical alarms faster.

Finally, in the fourth pilot (N=22, with n=5 in each of the low-frequency alarm conditions, and n=6 in each of the high-frequency alarm conditions), I added alarm frequency as a second between-subjects independent variable alongside display type. I now had a high- and low-frequency alarm condition crossed with the display condition (alarms only versus HWD-alarms). There were 45 alarms per block in the high-frequency condition and 30 in the low-frequency condition. Data from Pilot 4 suggested the participants in the HWD-alarms condition were still treating alarms faster, while leaving fewer untreated. Unexpectedly, there were no differences in treatment latency due to the interaction of alarm frequency and display condition.

The final design for Study 4 was identical to Pilot 4 with two between-subjects factors (display condition and alarm frequency), N=76, and n=19 in each condition. To further examine the potential differences when alarm frequencies were low versus high, I reduced the number of alarms in the low-frequency alarms condition from 30 to 22. The number of alarms in the high-frequency alarms condition remained unchanged at 45. For Study 4, I hypothesized that participants in the HWD-alarms condition would treat their clinical patient alarms faster, and more often, than participants in the alarms only condition. Additionally, I hypothesized that participants in the HWD-alarms condition would perform the ongoing task for longer than participants in the alarms only condition.

Regression analyses revealed that participants treated clinical alarms faster in the HWD-alarms condition, regardless of whether there was a high or low frequency of alarms (see top left of Figure 11). Participants using the HWD were also able to spend more time performing the ongoing dosage calculation task than did participants using alarms only, but only in the high alarms condition and not in the low alarms condition (see top right of Figure 11). Participants

using the HWD reported the same level of workload as participants relying on alarms only (see middle left of Figure 11). Participants using the HWD left fewer alarms untreated in both alarm frequency conditions (see middle right of Figure 11). Moreover, participants' accuracy on the ongoing dosage calculation questions was not affected across the different conditions (see bottom of Figure 11).

An additional regression was conducted to investigate differences in the management of clinical and sensor alarms across the display and alarm frequency conditions (Figure 12). This regression revealed that participants in the alarms only condition were, indeed, unable to differentiate between clinical and sensor alarms, as indicated by the nearly identical treatment latencies. Participants in the HWD-alarms condition, however, could identify which alarm was triggered in each room, allowing them to delay (or deny) their responses to the irrelevant sensor alarms.

The data from Study 4 suggest that the continuous information display on the HWD conveyed meaningful information to the participant, influencing their ability to prioritize clinically relevant alarms over nonactionable technical alarms. Participants without the HWD had to guess from which patient and room the clinical alarms were coming. If these outcomes were to generalise to a real multiple patient hospital ward setting, clinicians wearing HWDs may be able to make better treatment decisions when faced with alarms, without compromising their performance on the task at hand. In the next phase, my goal was to take another step towards clinical implementation, by testing whether nursing trainees could maintain awareness for specific changes to simulated patients while they performed a complex simulated patient assessment.

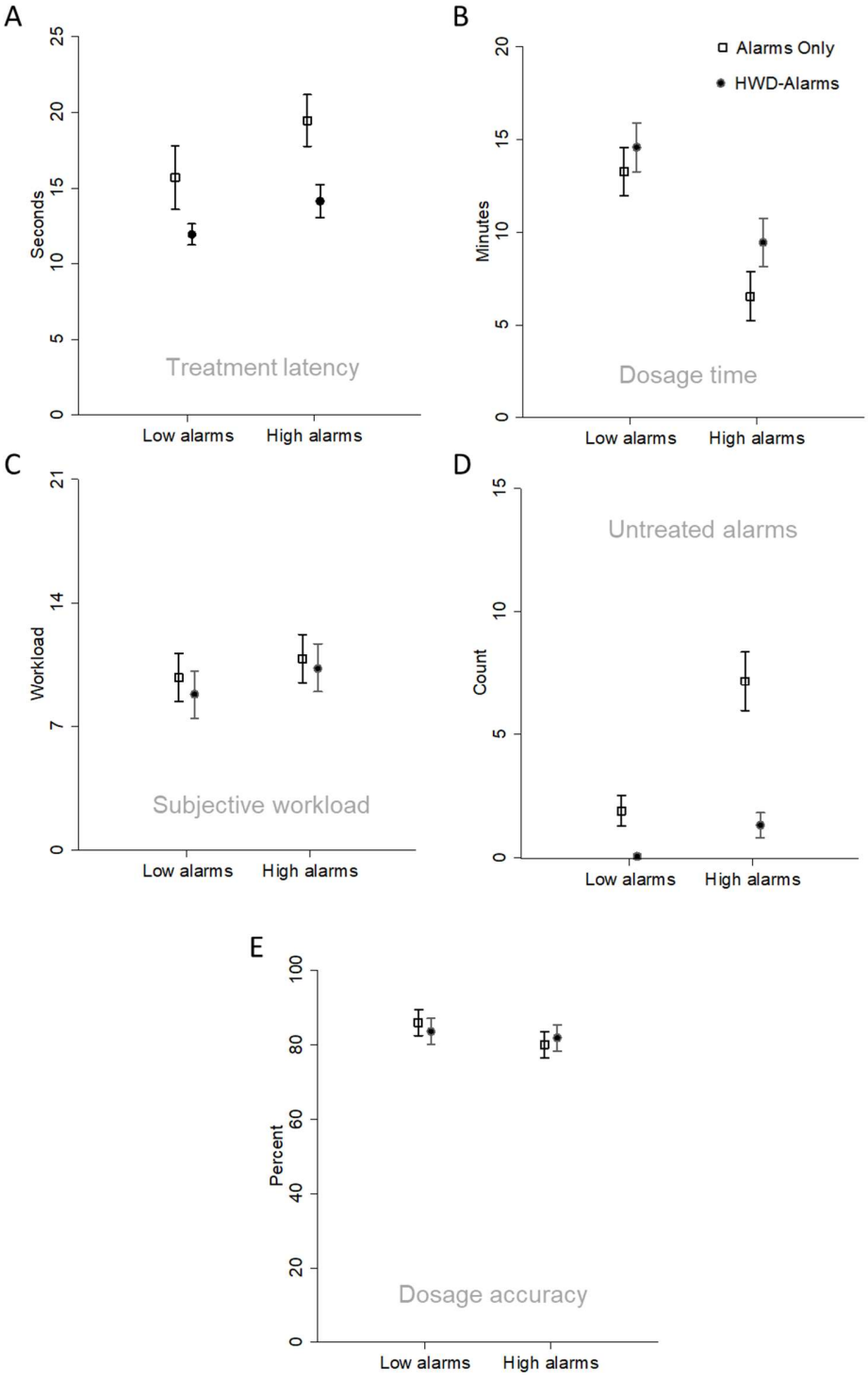

Figure 11. Results from regression analyses in Study 4. (A) observed values and 95% confidence intervals for treatment analysis (data were log transformed for analysis making predicted values less easy to interpret), (B)–(E) predicted values and 95% confidence intervals for dosage time, subjective workload, untreated alarms, and dosage accuracy.

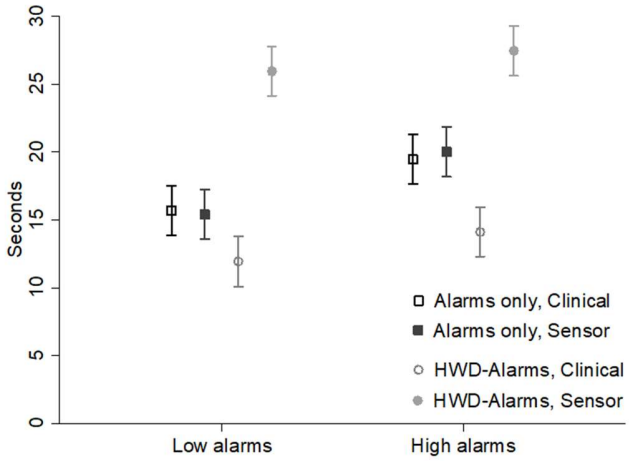

Figure 12. Predicted treatment latencies and 95% confidence intervals for clinical and sensor alarms across all conditions.

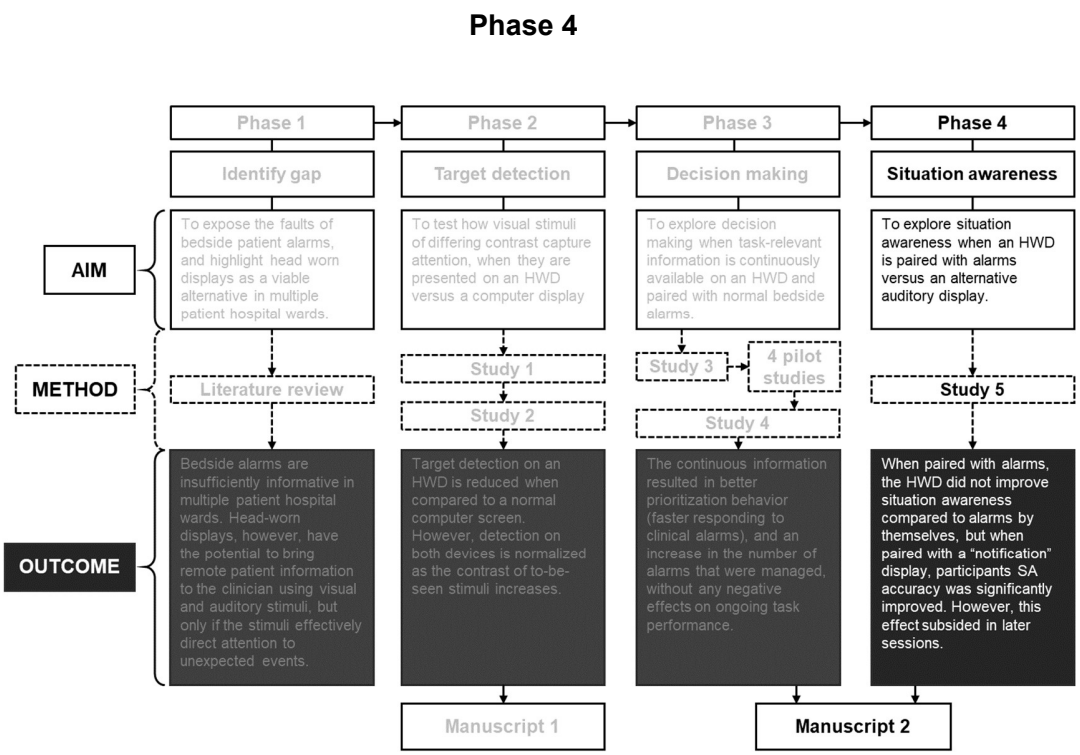

Figure 13. Phase 4 of the research program.

Study 5: Situation awareness, monitoring, and ongoing task performance with clinician trainee participants

Study 4 showed that information displayed on an HWD can be used to prioritize tasks in a simulated hospital environment. Specifically, participants using the HWD could check the display for clinically relevant patient alarms, and respond to them faster and more often. In most cases, however, clinicians do not have to move from room to room managing alarms. Instead, they spend longer periods of time with patients, or completing other tasks, and typically only manage alarms after those other responsibilities are complete. Thus, the behaviour seen in Study 4 is not entirely generalizable to a clinical context, but instead suggests that, in the event of a critical patient deterioration, the HWD might facilitate earlier recognition and prioritisation.

A more representative measure to test the effects of HWDs would be situation awareness, which has been associated with the quality of care being delivered by nurses (Sitterding, Broome, Everett, & Ebright, 2012; Stubbings, Chaboyer, & McMurray, 2012). Specifically, Stubbings et al. (2012) suggested that increases in nurses' situation awareness can improve decision-making and patient outcomes. Moreover, they discuss the need for curricula that directly address situation awareness during training. HWDs, however, may facilitate better situation

awareness even when training is limited. Having continual access to the real-time status of multiple patients, via an HWD, might improve clinicians' awareness of patient-related events, improving response times during deteriorations.

Integrated visual displays have been found to increase clinicians' situation awareness in intensive-care nursing (Koch et al., 2013) and anaesthesiology (Zhang et al., 2002). These displays have been fixed at the bedside rather than being head-worn but the findings, nonetheless, show that advanced display technologies can improve awareness of patient related events. Moreover, recent studies have tested advanced auditory displays against the current alarm standards. This body of research shows that more informative earcons (short auditory motifs) (Hickling et al., 2017) and speech messages (Li et al., 2017) can increase awareness for patients' status. Moreover, well-design auditory 'icons' can improve recognition and localisation compared to the current bedside alarm sounds and other alternatives (Edworthy et al., 2017). Thus, the next goal was to explore whether mobile HWDs together with advanced auditory displays would improve situation awareness (Endsley, 1995) in the multiple patient microworld. Study 5 was included as Experiment 2 in Manuscript 2 (Appendix B).

Study 5 was designed to compare participants' awareness of patient related events in three different display conditions: alarms only, HWD-alarms, and HWD-notifications. Awareness was operationalised using the situation awareness global assessment technique (SAGAT). In addition, I measured participants' performance on an ongoing, computerised patient assessment task as well as their subjective workload. Each participant experienced all three conditions over the course of three separate sessions (a within-subjects design).

In the alarms only condition, participants had to rely on bedside alarms. As in Study 3 and Study 4, the alarms were triggered by clinically relevant changes to a patient's status (exceeding thresholds) or by irrelevant changes in status (technical failures). The alarms could take on one of two sounds (IEC60601-1-8, 2005– 08). The first was a "warning" alarm (IEC General medium) that indicated that a vital sign was too high or too low. The second was a "critical" alarm (IEC General high) that indicated that a vital sign was critically high or critically low. To get any further information about the alarm, the participant had to physically move to the hospital ward computer, and then into one of the patient's rooms using a mouse click.

In the HWD-alarms condition, participants heard bedside alarms, but they could also use the HWD to see the vital signs of all six patients. Access to the patient's vital signs meant that they did not have to physically move to the hospital ward computer to update their awareness of their patients' status.

In the HWD-notifications condition, participants used the HWD as in the HWD-alarms condition. Instead of hearing alarms, however, participants heard an intermittent, cycling, auditory

sequence display that played short notification sounds that conveyed the specific status of each patient in order (see Figure 14).

The sequence display was triggered whenever the status of a patient changed, or after 30 seconds (see Figure 15). There were three sounds associated with three potential patient states. The first was a “boop” sound that was used when the patient had no abnormal vital signs. The second was a higher pitched “warning” beep, with vibrato (tremolo), that was used when the patient had at least one too-high or too-low vital sign. The third was an even higher-pitched “critical” beep with faster vibrato (tremolo), indicating that at least one of the patient’s vital signs was critically high or low. One sound was played for each patient depending on their respective vital signs, with a 400 millisecond pause between patients.

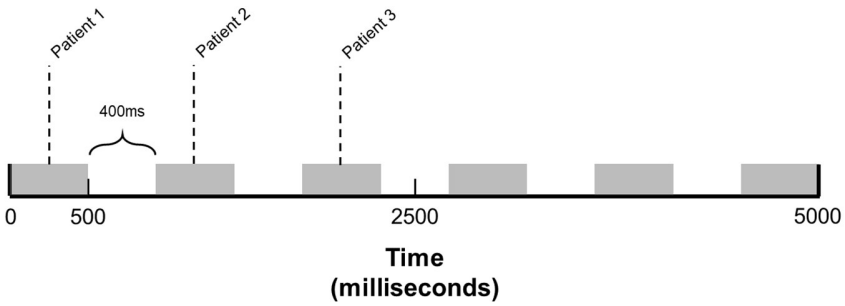

Figure 14. Diagram of the sequence display used to convey the status of all six patients. The grey boxes indicate when the auditory stimuli were being played. Each patient’s earcon was played for 500 ms followed by 400 ms of silence before the next patient. The total sequence lasted 5 seconds.

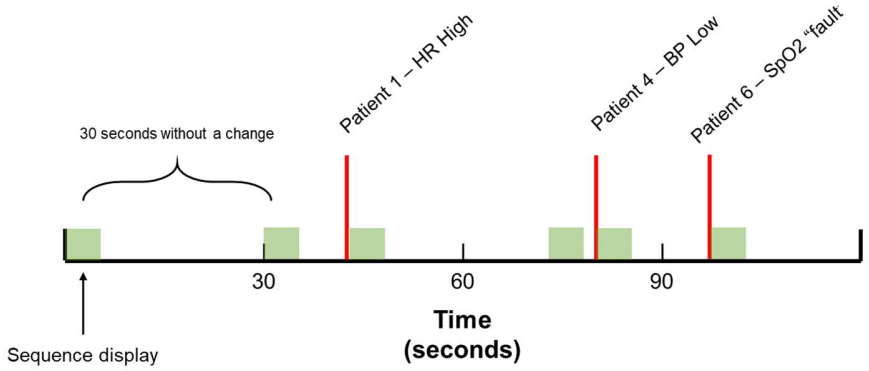

Figure 15. Diagram of the pattern, over time, of the sequence display. Green boxes indicate when the auditory stimuli were being played. Red lines indicate unexpected patient events. Unless triggered by a patient deterioration, the sequence display played every 30 seconds.

The ranges I used for the vital signs for each patient were taken from the Queensland Adult Deterioration Detection System (Q-ADDS). Each vital sign had its own specific range of values that indicated when the vital signs was normal, too high, or too low. Figure 16 shows the ranges for heart rate, blood pressure, and oxygen saturation. These ranges were used as guides for the alarms and the notifications. The normal ranges are depicted in white. Warning alarms/notifications were triggered when the value for a vital sign was crossed from white into the yellow/orange range. Critical alarms/notifications were triggered when the value for a vital sign crossed from yellow or orange into the red/purple range.

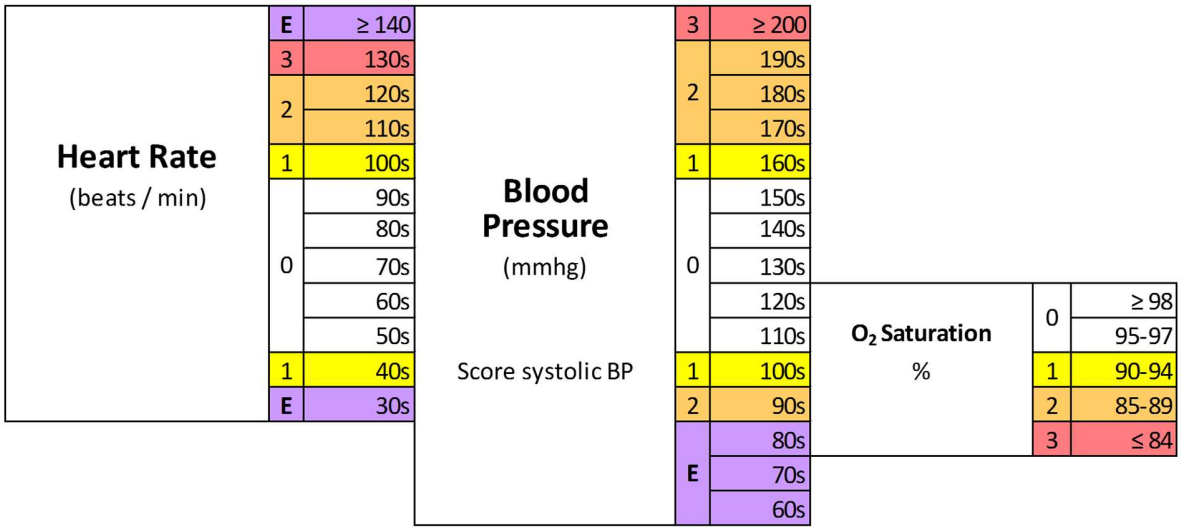

Figure 16. Diagram of values for heart rate, blood pressure, and oxygen saturation from the Q-ADDS.

Given the simplicity of the previous experiments, with respects to clinical patient monitoring, I aimed to increase the representativeness of Study 5 by making changes to the population and the tasks. For this study, I recruited second and third (final) year nursing students who had background knowledge of patient physiology and patient monitoring, as well as direct experience performing patient monitoring tasks in hospital. At The University of Queensland, second and third year students spend the majority of their week at the hospital. For second year students, three quarters of their time is spent in a classroom, and a quarter of their time is spent providing care to patients. For third year students, those responsibilities are switched. All students were familiar with the Q-ADDS and each had completed multiple patient assessments in hospital.

For the ongoing task, participants completed computerized patient assessments (using vSim™ for Nursing, Laerdal) that required them to choose (using mouse clicks) from a series of actions, ranging from washing hands to administering CPR. Each computerized patient in the

assessment had a different background, physiology, set of symptoms, and deterioration schedule, that made each scenario unique. The participants' goal was to complete as much of the patient assessment (one scenario per block) as they could, as accurately as they could, while simultaneously monitoring the six simulated patients for changes. Unlike Study 3 and Study 4, participants were now just monitoring for changes to a patient's status and were not required to intervene by pressing the "Fix sensor" or "Treat patient" buttons.

Periodically, the patient monitoring task would freeze, prompting the participant to answer two questions relating directly to the status of the patients being monitored remotely in one of the three display conditions; for example, "Which patient has critically low blood pressure?" The questions were designed to be as specific as possible, requiring information from the visual display, but three of the ten questions could have still been answered using the notification display alone. Table 3 contains an example set of questions from a single block. All blocks used a similar set of questions, but with the specific patient, vital sign, and order changed to reflect the different scenario of events for each patient in each unique block. After the participants responded to the questions, the simulation would resume from where it left off.

Table 3

Example questions from one of the blocks in Study 5

Question Number	Question
1	Which patient has the lowest SpO2?
2	How many of Patient 5's vital signs are outside the normal range?
3	Which patient has the highest HR?
4	How many patients have BP outside the normal range?
5	Which of Patient 1's vital signs is outside the normal range?
6*	Did the status of Patient 2 change since the last freeze?
7*	Which Patient's BP just dropped to 0?
8	Which patient has the highest BP?
9*	How many patients have completely normal vital signs?
10	How many of Patient 2's vital signs are outside the normal range?

*Note. Questions marked with * could have been answered with the notification display alone.*

My intention was to recruit at least 20 nursing trainees for Study 5, each of whom would complete all three display conditions over the course of three sessions. I approached a total of six universities, but only three approved my request to advertise. Of the three universities, advertising was only successful with one, resulting in a total sample of 13 participants.

The results are reported in full in Manuscript 2 (Appendix B). A mixed effects linear regression revealed that when participants were using the HWD and notification display, they answered more situation awareness questions correctly than when they were using alarms only or HWD and alarms, but only in session 1. Interestingly, there was not a similar improvement on

the SAGAT questions between the alarms only and the HWD-alarms condition (see top left of Figure 17).

Additional regression analyses revealed a significant improvement in scores on the ongoing patient assessment task for the HWD-alarms and the HWD plus notification conditions, compared to the alarms only condition (see top right of Figure 17). The improved assessment scores in the HWD conditions are most probably related to the participants' ability to quickly obtain information about other patients from the HWD, without having to completely stop the patient assessment task and move to the microworld computer. A further analysis showed that subjective workload was reduced in both HWD conditions compared to the alarms only condition, across all three sessions (see bottom left of Figure 17). A final regression revealed that participants found the alarms in the alarms only condition to be "more annoying" than the alarms in the HWD-alarms condition, and the notifications in the HWD-notifications condition (see bottom right of Figure 17).

Altogether, the results of Study 5 show that HWDs can be used to improve performance, primarily when paired with an informative auditory display that alerts participants to changes more effectively. When using the HWD plus notification display in the first session, participants answered situation awareness questions more accurately than when using alarms only or HWD-alarms. In our paradigm, however, this benefit shrank as experience with the tasks increased. With an ongoing task that limited the participants' ability to freely stop and start the task, I expect that SAGAT accuracy in the alarms only condition would not have improved in later sessions. Additionally, the HWD was associated with improvements in the ongoing patient assessment task, which may have been because participants did not have to spend as much time checking on the other patients in the hospital microworld.

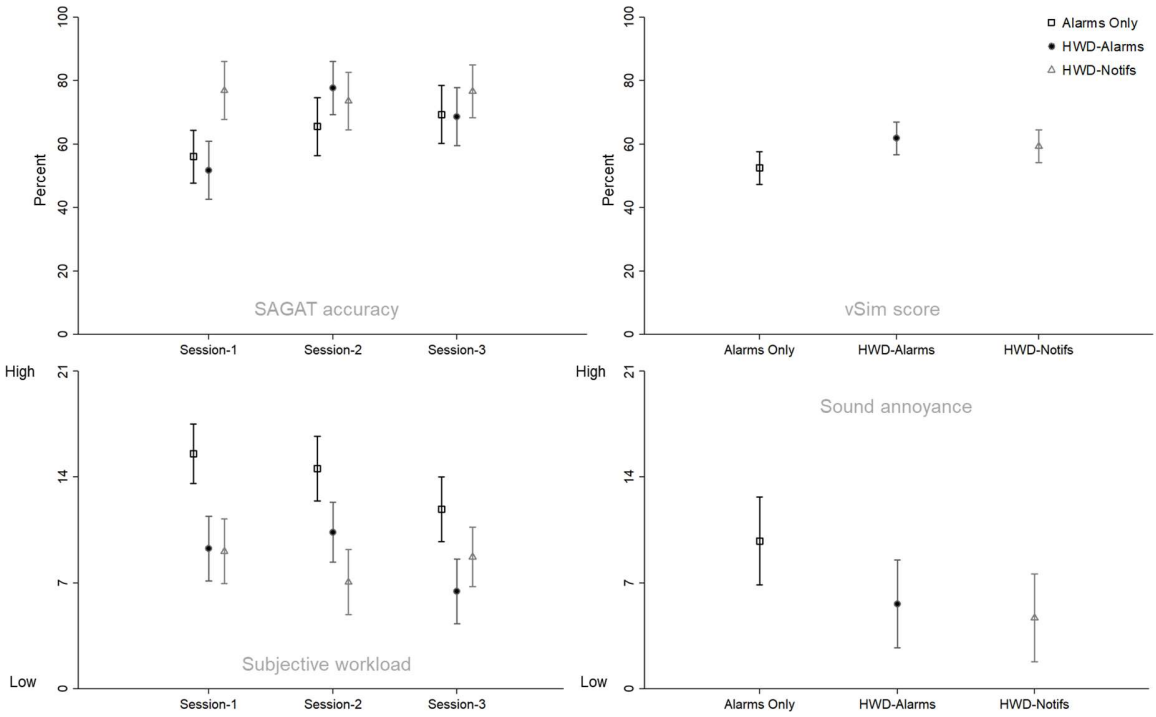

Figure 17. Predictive models with 95% confidence intervals produced from the regression analyses in Study 5.

Discussion

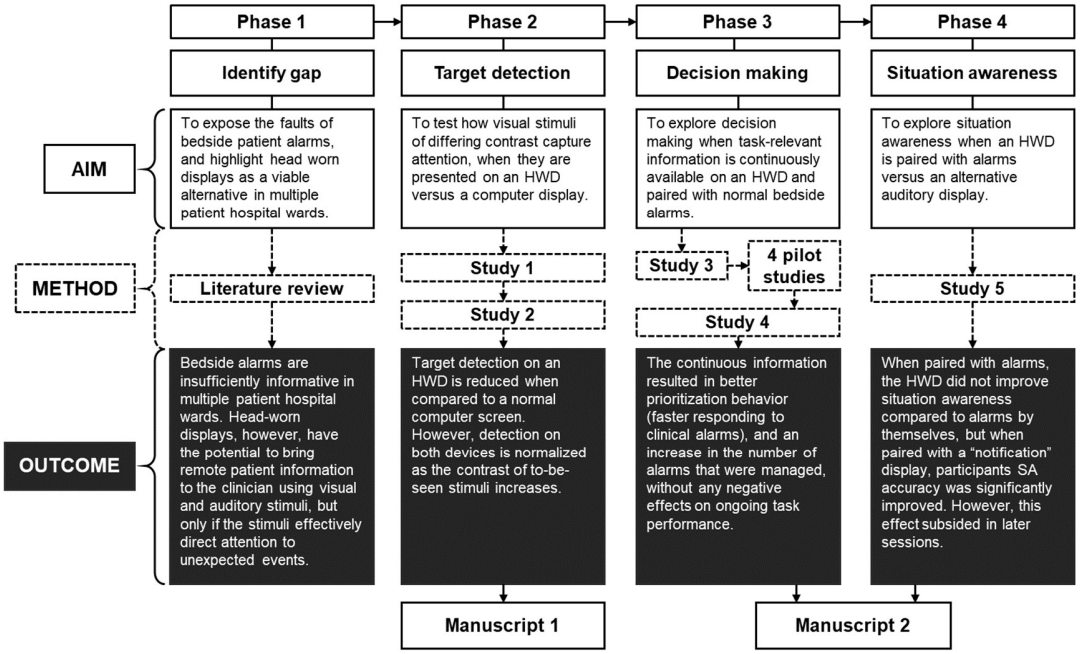

Figure 18. Overview of program of research

The purpose of this thesis was to explore how HWDs and continuous information displays could be used to guide a viewer's attention to clinically relevant patient alarms to increase their awareness for the status of multiple patients. Current bedside monitoring equipment and patient alarm systems in hospitals are designed to inform clinical staff of patient-related events that may require attention, but the alarms occur frequently and often do not require an immediate response from the clinical staff. This generates distrust in the alarm, which is associated with high rates of delayed or inappropriate responses to alarms. With HWDs that display continuous information, however, clinicians may be able to continually update their awareness of their patients' status, and use the display to make better decisions regarding when to move to a patient's bedside. In the sections that follow, I summarise the findings, place them in theoretical and practical context, and discuss limitations and directions for future research.

Implications of detecting peripheral stimuli on HWD

In Study 1 and Study 2 within Phase 2, I tested participants' ability to detect abrupt changes to peripheral stimuli displayed on an HWD. Compared to target detection rates on a computer screen, detection rates for stimuli on the HWD were reduced (Manuscript 1 in Appendix A). These data corroborate the findings from Yantis and Jonides (1990) and Nikolic et al. (2004)

regarding detection rates for abrupt onset and peripheral stimuli, respectively. However, the data extend these findings to peripheral stimuli displayed in a constant location on an HWD. Moreover, on an HWD, the reductions in target detection due to peripheral displacement and lack of distinctiveness are more pronounced than on a computer monitor. For target stimuli closer to the direction of gaze and with strong distinctiveness, the difference in detectability between the HWD and computer monitor was lessened, but target change detection on the HWD was still significantly worse than on the computer display.

These studies suggest that capturing visual attention using a monocular HWD may be more difficult than when stimuli are viewed by both eyes on a computer monitor, probably due to the monocular presentation of the stimuli. Stimuli seen by only one eye on an HWD are subject to a variety of perceptual phenomena including binocular rivalry (Winterbottom, Patterson, Pierce, Covas, et al., 2006), suppression related to unmatched imagery from both eyes (Arnold, 2011), depth of focus limitations (Winterbottom et al., 2007) and the Gestalt concept of common fate (Koffka, 1935). No matter the cause, the data show that visual stimuli may not be sufficient for reliably attracting attention to important stimuli on a monocular HWD.

The data in in Phase 2 suggest several design recommendations concerning how stimuli should be displayed on an HWD to successfully, and reliably, capture attention exogenously. Specifically, transient stimuli that need immediate attention should be displayed as distinctively as possible from the background. Doing so will increase the likelihood that the stimuli will be detected, but as the HWD user's gaze is directed away further from the HWD, thereby increasing the distance between fovea and HWD, the likelihood of detecting even distinctive stimuli decreases. A better solution is to use simultaneous auditory displays that are redundant with the visual display, as in Woodham et al. (2016).

In many contexts, however, important task-relevant stimuli would not be displayed for 200 ms, and will not rely on immediate, exogenous attention capture. For stimuli that change over the course of minutes, and depending on the goals of the perceiver, endogenous attention will probably drive display checking behaviour. Still, in the event of a critical change, and when auditory displays may not be an option, designers should maximize the likelihood of capturing attention using the HWD display by using stimuli that contain the highest possible brightness, and that appear to be in motion.

Implications of HWDs on alarm prioritisation behaviour

In Studies 3 and 4 within Phase 3, I tested participants' ability to use continuous information to prioritise relevant alarms over irrelevant alarms. Much like previous research on continuous "monitor watchers" (Funk et al., 1997; Watkins et al., 2016), I found that participants using the

continuous information displays responded to clinically relevant alarms faster, and more often than participants who had to rely on alarms alone for patient information (Experiment 1, Manuscript 2 in Appendix B). This result was found regardless of how frequently alarms occurred. Additionally, when an ongoing task was introduced in this paradigm, participants using the HWD performed the ongoing task for longer, and the advantage was particularly apparent during scenarios with a high frequency of alarms.

Furthermore, Studies 3 and 4 partially replicate the ineffectiveness of bedside alarms for guiding attention to relevant patient events. Although highly simplified, the microworld showed that alarms alone provided insufficient information for participants to make informed decisions about which patients to prioritise in the event of multiple alarms. Thus, in hospital wards where alarms occur frequently and are often unreliable (Paine et al., 2016; Ruskin & Hueske-Kraus, 2015), and when the average response time to alarms can exceed seven minutes (Bonafide et al., 2017), an HWD may improve patient prioritisation for any relevant, actionable alarms that do occur. Faster recognition and improved prioritisation might result in improved patient outcomes if deteriorating patients get treatment faster, before complications can arise.

Implications of HWDs on situation awareness

In the final study, Study 5, I tested participants' ability to use HWDs to maintain situation awareness of the status of multiple patients. I found that the HWD did not universally improve situation awareness (Experiment 2, Manuscript 2 in Appendix B). Specifically, when the HWD was used as a supplement to normal bedside alarms, the participants' responses to the situation awareness questions were no more accurate than their responses when they had to rely on the alarms by themselves. However, when the HWD was combined with the cycling display of individual patient notifications, participants' responses to the situation awareness questions were significantly more accurate than when they had to rely on alarms, or even alarms and an HWD. This effect, however, was only present in the first session suggesting that participants situation awareness improved in our paradigm regardless of the display condition. Therefore, in subsequent sessions, accuracy in the alarms only condition and HWD-alarms conditions were not significantly different from the HWD-notifications condition.

These data are supported by previous findings regarding improved situation awareness using integrated visual displays during single patient monitoring (Koch et al., 2013; Zhang et al., 2002), but also show that combining advanced visual displays with standard bedside alarms may not always result in better patient outcomes. The differences in SAGAT accuracy between HWD conditions is explained by the lack of information coming from the standard alarms. The notification display played sounds for all patients immediately when a change had occurred,

directing attention to the display at the exact moment, or within 5 seconds, that something new happened, as with exogenous attention. The alarms, however, were not able to consistently cue the participant to look at the display when important changes were happening. The very first instance of the alarm sound was paired with the patient's change in status, but subsequent instances of the alarm were simply reminders. When more than one alarm was sounding, it was much more difficult to differentiate an old alarm sound from a new alarm sound, which made acquiring information from the visual display less reliable, when using the alarms as exogenous cues. Instead, checking the display in the HWD-alarms condition was probably driven by top-down, endogenous goals. In fact, this interpretation might help explain the lack of differences in the second and third sessions, when endogenous attention within the microworld was more developed. Whatever the reasons may be, the differences in SAGAT accuracy in session 1 corroborate Edworthy et al. (2017) showing that latching bedside alarms provide very little context about which patient is potentially in distress, and what the nature of that distress is.

A further implication of Study 5 is that the inverting of colours of the vital signs presented on the HWD may not have been distinctive enough to capture attention reliably (as in Study 2) without the addition of an auditory stimulus. Each change in a patient's status was coupled with an immediate change in the contrast of the values displayed on the HWD. Had these changes in contrast been powerful enough to attract attention to the display, the improved accuracy would have been present in both HWD conditions, regardless of the kind of auditory display used. The lack of attention capture by the visual stimuli in Study 5 extend the original findings from Study 2 further. Specifically, detecting visual stimuli on an HWD, even those high in contrast, may be severely limited when focused attention is directed elsewhere, as with perceptual loading (Lavie, 2005) and inattentional blindness (Simons & Chabris, 1999). This suggestion is not novel (Browning, 2016; Dixon et al., 2013; Liu, Jenkins, Sanderson, et al., 2009; Tang, 2016), but provides further evidence that the advanced auditory display played an integral role in successfully guiding attention to meaningful information on the HWD.

A final important finding was the effect of HWD use on the patient assessment task. In both HWD conditions, participants obtained significantly higher scores on the patient assessment than when they had to rely on alarms alone. Previous research would suggest that there are dual-task costs associated with the use of HWDs (He et al., 2015; Mustonen, Berg, Kaistinen, Kawai, & Hakkinen, 2013; Sawyer et al., 2014; Woodham et al., 2016); however, in Study 5, the HWD was associated with improvements in ongoing task performance. This result may have been related to the time costs associated with interrupting the patient assessment to physically move to the other patients in the alarms only condition—costs that would exist in many clinical contexts. Interruptions have been widely considered to be a hazard in healthcare (Grundgeiger &

Sanderson, 2009; Liu, Grundgeiger, Sanderson, Jenkins, & Leane, 2009), but the impact of HWDs on interruptions, or vice versa, has not yet been determined. Still, if improvements to ongoing task performance generalise to a clinical monitoring context, a clinician using an HWD to maintain awareness of his or her patients may complete their tasks faster while wearing the HWD, or make fewer clinical errors, for example forgetting a step in a procedure.

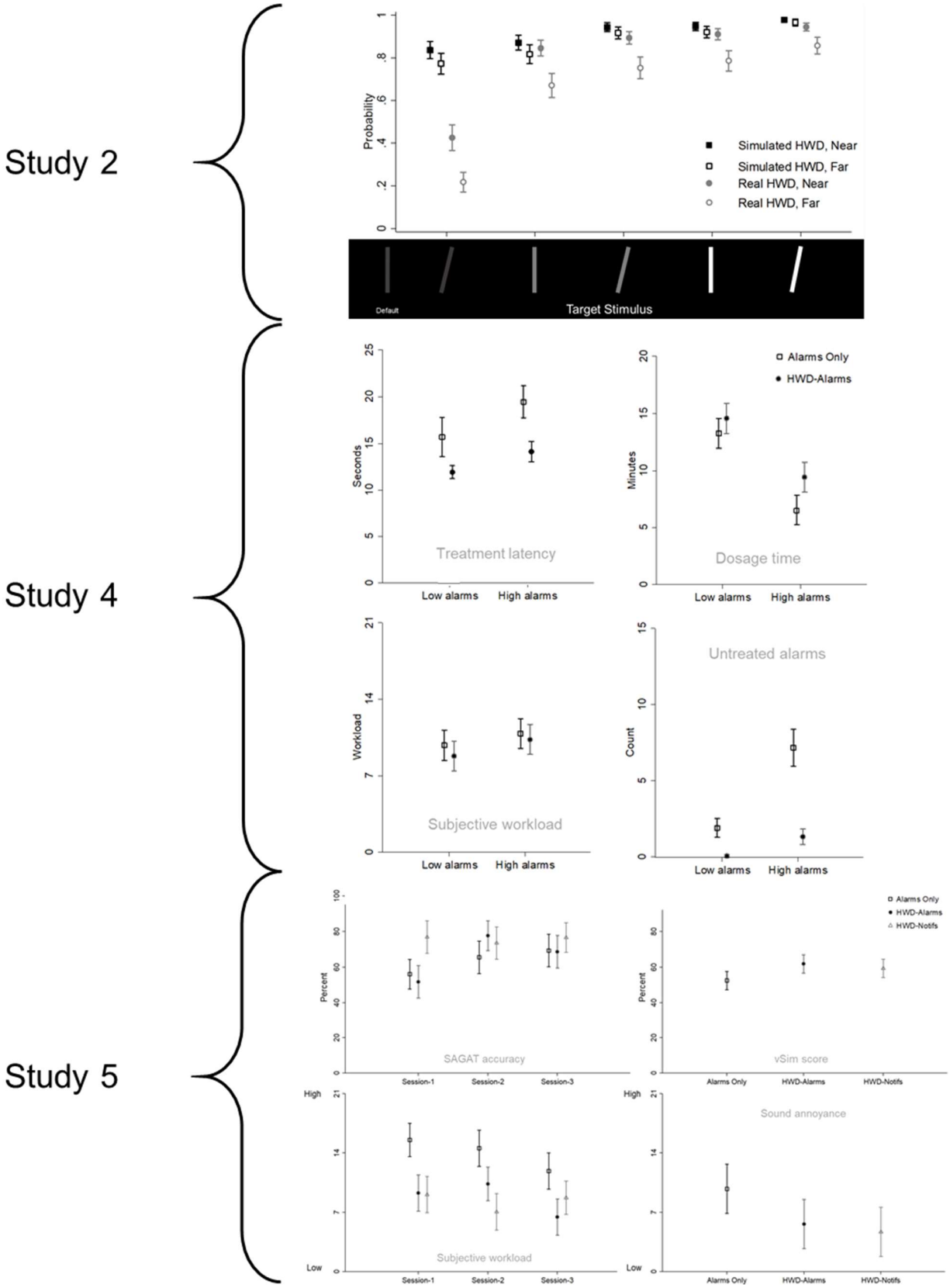

Figure 19. Summary of the data from each of the studies investigating HWDs.

Summary of findings

The program of research discussed in this thesis offers several guidelines for HWD use (Figure 19). For simple, visual notifications on an HWD, stimuli need to be as distinctive as possible, but even these stimuli are likely to be missed if they are far into the periphery (Study 2) or not accompanied by a redundant audible cue (Study 5). Continuously available information from multiple remote sources can be used by participants to make prioritisation decisions in environments with both high and low alarm frequencies, without affecting their performance on simple cognitive tasks (Study 4). Using an HWD can reduce the effects of having to briefly stop one task to prioritise another, resulting in more uninterrupted time on ongoing tasks (Study 4) and, in some instances, better performance (Study 5). HWDs can be an effective means for maintaining awareness of the status of multiple remote sources, but only with auditory displays that are better equipped for guiding attention to important events (Study 5). When faced with a complex patient assessment task, participants using HWDs reported lower workload than participants relying on just alarms (Study 5). Similarly, participants using HWDs described the auditory displays, even in the HWD-alarms condition, as being less annoying than the alarms in the alarms only condition (Study 5).

Limitations and future research

The current program of research has shown the potential for HWDs to be used in clinical contexts, extending the positive findings from single patient studies (Liu, Jenkins, Sanderson, Fabian, & Russell, 2010; Vorraber et al., 2014) to multiple patients. Specifically, the results of the current studies suggest that HWDs might be potential mediators to the alarm problem (Woods, 1995) by decentralizing patient information, which can contribute to better decision-making and improved situation awareness without negatively affecting performance on other tasks. However, to avoid perceptual limitations imposed by HWDs (Pascale, Sanderson, Liu, Mohamed, Stigter, et al., 2015; Patterson, Winterbottom, & Pierce, 2006) and task loading (Lavie, 2005), the visual stimuli should be paired with advanced auditory stimuli (Edworthy et al., 2017; Paterson et al., 2017) that guide attention to the visual display meaningfully. Still, before HWDs can be put into practice for multiple patient monitoring, future research will need to address several questions relating to sustained use (vigilance), the benefits in more representativeness testing contexts, display checking behaviours, and the ease of access to information using different integrative displays.

The first concern is that HWDs have not been tested over periods of time equivalent to the duration of a clinical shift. Some shifts can last longer than 12 hours, which will inevitably influence HWD use. On the one hand, it is unknown how fatigue affects participants' ability to use

an HWD effectively. Moreover, increased fatigue may worsen a participant's ongoing task performance when their attention is already being strained by extended shifts. On the other hand, prolonged experience with the HWD may encourage the development of different, goal-driven strategies for monitoring patients using an HWD. Development of endogenous strategies, may reduce the effects of divided attention, and potentially ease some of the strain during extended shifts. Similarly, as with any modern technology, more experience using the device will inevitably require less focused attention, making it easier to use. Testing HWDs over representative time periods, however, may not be feasible in a laboratory study, and thus might require trials in the environments of interest.

A further branch of this research needs to use more representative environments and participants. In the current studies, patient deteriorations only approximated those seen in hospitals. I designed the scenarios in each study to address certain questions, but a clinical context will not be so highly controlled and simplified. Actual patient deteriorations are more complex and involve combinations of vital sign indicators rather than a single vital sign increasing or decreasing away from normality. Similarly, the ongoing tasks were simplified, particularly with regards to the dosage calculation task In Study 4. The vSim for Nursing patient assessment task, in Study 5, was complex and required focused attention, but performing actions like administering a drug manually, will require further increases in focused attention greater than that of making button presses. Differing demands on attention will inevitably change how well the device can be used. Additionally, I found improvements to prioritisation behaviour and situation awareness in separate studies, but a future study should investigate them, together, in a single experiment.

Future studies should also investigate display checking behaviours. For example, an eye-tracker might have helped explain the data further, had it been an option, particularly in Study 5. It would have been useful to compare how many saccades were being made to the HWD in the two HWD conditions. The worsened accuracy in the first session in the HWD alarms condition may have been the result of fewer saccades, because participants didn't have appropriate cues. Alternatively, they might have sampled the display an equal number of times, but at times when the information was less meaningful. A recent study that used eye-tracking with an HWD suggests that participants have to check the display frequently to acquire information when there are not exogenous cues to guide them (Schlosser, Sanderson, Grundgeiger, Liu, & Loeb, 2016). Schlosser found, however, that frequent checking still resulted in slower event detection and worsened performance on ongoing tasks compared to a computer display condition that utilised distinct flashes as cues towards important events. In my Study 5, it is unknown whether the displays, both visual and auditory, contributed to more checking or less checking.

Of considerable interest, and warranting further investigation, is the difference found in Study 5, where participants' situation awareness in the HWD-alarms condition, operationalised by SAGAT accuracy, was significantly lower than participants' situation awareness in the HWD-notifications condition. This effect was eliminated after the first session in our paradigm, potentially due to familiarisation with the specific tasks and carryover effects across conditions, but the effect will possibly be more prominent, and more persistent, under constrained conditions that approximate clinical contexts.

Finally, the visual and notification displays used in the current studies were underdeveloped. They contained the basic information necessary for participants to update their mental model of the environment, but there is room for improvement. Multimodal displays using vision, audition, and touch can interact in a multitude of ways to generate an information rich environment (Sanderson, 2006) above that of visual displays of numbers. HWDs with trend information may enhance the detectability of deteriorations that occur over periods of time. HWDs with configurable displays may make information acquisition faster, possibly reducing the amount of attention necessary to get an update (Koch et al., 2013; Zhang et al., 2002). With respects to auditory displays, the studies investigating spearcons and earcons (Edworthy et al., 2017; Hickling et al., 2017; Paterson et al., 2017) might uncover auditory displays that could be used to guide attention to the display with greater reliability. Some of these displays are even rich enough to convey specific patient information without requiring an HWD, but information becomes more difficult to retain as the number of patients being presented increases (Hickling et al., 2017). That effect, however, might be less prominent when advanced auditory displays are paired with visual stimuli (Bell, 2017).

Conclusion

The purpose of this thesis was to investigate the attention capturing, and attention guiding effects of HWDs. My research questions were driven by previous studies investigating unreliable and uninformative alarm systems in hospitals that often contribute to unintentional patient mismanagement. HWDs, however, could be used in place of alarms to facilitate greater awareness of patients' wellbeing by providing continuous access to information that would otherwise be inaccessible without physically moving to a centralized monitoring display, or to an individual patient's bedside.

In five studies, I tested whether HWDs effectively captured visual attention, whether continuous information improved patient prioritisation, and whether the continuous information increased situation awareness. I also tested whether the combination of HWD and an alternative auditory notifications display would result in increased situation awareness, compared to an

HWD and alarms. The results were encouraging, suggesting that HWDs may improve patient prioritisation, and increase situation awareness. However, the visual display of the HWD should not be relied on for conveying important information due to perceptual constraints that limit the attention guiding effects of visual stimuli. Instead, visual stimuli should be paired with auditory displays that guide attention to the display when important changes are occurring, like my notification display. How visual and auditory displays interact will require further research; but it is safe to assume that some combination of the two will significantly improve the current state of patient monitoring, eliminate reliance on uninformative bedside alarms, and improve patient outcomes.

Bibliography

Arnold, D. (2011). I agree: binocular rivalry stimuli are common but rivalry is not. *Frontiers in Human Neuroscience, 5*(157). doi:10.3389/fnhum.2011.00157

Baig, M. M., GholamHosseini, H., & Lindén, M. (2015). *Tablet-based patient monitoring and decision support systems in hospital care.* Paper presented at the Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE.

Bell, E. (2017). *Time-compressed speech: Spearcon use for monitoring multiple patients.* (BPsySci(Hons)), The University of Queensland, St Lucia, Australia.

Block III, F. E., & Block Jr, F. E. (2015). Decreasing False Alarms by Obtaining the Best Signal and Minimizing Artifact from Physiological Sensors. *Biomedical Instrumentation and Technology, 49*(6), 423-431.

Bonafide, C. P., Localio, A. R., Holmes, J. H., Nadkarni, V. M., Stemler, S., MacMurchy, M., . . . Keren, R. (2017). Video Analysis of Factors Associated With Response Time to Physiologic Monitor Alarms in a Children's Hospital. *JAMA Pediatr, 171*(6), 524-531. doi:10.1001/jamapediatrics.2016.5123

Browning, C. (2016). *Head Worn Displays with Visual Cues and Animal Call Alerts for Multiple Patient Monitoring.* (BPsySci(Hons)), The University of Queensland, St Lucia, Australia.

Büttner, S., Funk, M., Sand, O., & Röcker, C. (2016). *Using head-mounted displays and in-situ projection for assistive systems: A comparison.* Paper presented at the Proceedings of the 9th ACM international conference on pervasive technologies related to assistive environments.

Cvach, M. M. (2012). Monitor alarm fatigue: an integrative review. *Biomedical Instrumentation and Technology, 46*(4), 268-277. doi:10.2345/0899-8205-46.4.268

Cvach, M. M., Frank, R. J., Doyle, P., & Stevens, Z. K. (2014). Use of pagers with an alarm escalation system to reduce cardiac monitor alarm signals. *Journal of Nursing Care Quality, 29*(1), 9-18.

Davidson, J. E., Agan, D. L., Chakedis, S., & Skrobik, Y. (2015). Workplace blame and related concepts: an analysis of three case studies. *CHEST Journal, 148*(2), 543-549.

Dixon, B. J., Daly, M. J., Chan, H., Vescan, A. D., Witterick, I. J., & Irish, J. C. (2013). Surgeons blinded by enhanced navigation: the effect of augmented reality on attention. *Surgical Endoscopy, 27*(2), 454-461. doi:10.1007/s00464-012-2457-3

Drake-Brockman, T. F., Datta, A., & Ungern-Sternberg, B. S. (2016). Patient monitoring with Google Glass: a pilot study of a novel monitoring technology. *Pediatric Anesthesia, 26*(5), 539-546.

Edworthy, J., Loxley, S., & Dennis, I. (1991). Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *Human Factors, 33*(2), 205-231.

Edworthy, J., Reid, S., McDougall, S., Edworthy, J., Hall, S., Bennett, D., . . . Pye, E. (2017). The Recognizability and Localizability of Auditory Alarms: Setting Global Medical Device Standards. *Human Factors*, 0018720817712004.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32-64.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology-Human Perception and Performance, 18*(4), 1030-1044. doi:Doi 10.1037//0096-1523.18.4.1030

Folk, C. L., Remington, R. W., & Wright, J. H. (1994). The structure of attentional control: contingent attentional capture by apparent motion, abrupt onset, and color. *Journal of Experimental Psychology: Human Perception and Performance, 20*(2), 317-329.

Funk, M., Parkosewich, J., Johnson, C., & Stukshis, I. (1997). Effect of dedicated monitor watchers on patients' outcomes. *American Journal of Critical Care, 6*(4), 318-323.

Gorges, M., Markewitz, B. A., & Westenskow, D. R. (2009). Improving alarm performance in the medical intensive care unit using delays and clinical context. *Anesthesia and Analgesia, 108*(5), 1546-1552. doi:10.1213/ane.0b013e31819bdfbb

Grundgeiger, T., & Sanderson, P. (2009). Interruptions in healthcare: theoretical views. *International Journal of Medical Informatics, 78*(5), 293-307.

He, J., Ellis, J., Choi, W., & Wang, P. (2015). *Driving while reading using Google glass versus using a smart phone: which is more distracting to driving performance?* Paper presented at the Proceedings of the Eighth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.

Hickling, A., Brecknell, B., Loeb, R. G., & Sanderson, P. (2017). Using a sequence of earcons to monitor multiple simulated patients. *Human Factors, 59*(2), 268-288.

Hinckfuss, K., Sanderson, P., Loeb, R. G., Liley, H. G., & Liu, D. (2016). Novel pulse oximetry sonifications for neonatal oxygen saturation monitoring: A laboratory study. *Human Factors, 58*(2), 344-359.

Horrey, W. J., & Wickens, C. D. (2004). Driving and side task performance: the effects of display clutter, separation, and modality. *Human Factors, 46*(4), 611-624.

IEC60601-1-8, I. I. S. (2005– 08). Medical electrical equipment, Part 1– 8: general requirements for safety— collateral standard: general requirements, tests and guidance for alarm

systems in medical electrical equipment and medical electrical systems. Geneva, Switzerland: International Electrotechnical Commission.

Imhoff, M., Kuhls, S., Gather, U., & Fried, R. (2009). Smart alarms from medical devices in the OR and ICU. *Best Practice & Research: Clinical Anaesthesiology, 23*(1), 39-50. doi:10.1016/j.bpa.2008.07.008

Jonides, J. (1981). Voluntary vs. automatic control over the mind's eye's movement. In J. B. Long & A. D. Baddelay (Eds.), *Attention and Performance IX* (pp. 187-203). Hillside, NJ: Erlbaum.

Knight, H. M., Gajendragadkar, P. R., & Bokhari, A. (2015). Wearable technology: using Google Glass as a teaching tool. *BMJ Case Reports, 2015*. doi:10.1136/bcr-2014-208768

Koch, S. H., Weir, C., Westenskow, D., Gondan, M., Agutter, J., Haar, M., . . . Staggers, N. (2013). Evaluation of the effect of information integration in displays for ICU nurses on situation awareness and task completion time: a prospective randomized controlled study. *International Journal of Medical Informatics, 82*(8), 665-675.

Koffka, K. (1935). *Principles of Gestalt psychology* (Vol. 44): Routledge.

LaFleur, T., Draper, M. H., & Ruff, H. A. (2001). Evaluation of Eye-Dominance Effects on Target-Acquisition Tasks Using a Head-Coupled Monocular Hmd. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 45*(18), 1433-1433.

Lavie, N. (2005). Distracted and confused?: selective attention under load. *Trends in Cognitive Sciences, 9*(2), 75-82. doi:10.1016/j.tics.2004.12.004

Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current Directions in Psychological Science, 19*(3), 143-148. doi:Doi 10.1177/0963721410370295

Lavie, N., Beck, D. M., & Konstantinou, N. (2014). Blinded by the load: attention, awareness and the role of perceptual load. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 369*(1641), 20130205. doi:10.1098/rstb.2013.0205

Li, S. Y., Magrabi, F., & Coiera, E. (2011). A systematic review of the psychological literature on interruption and its patient safety implications. *Journal of the American Medical Informatics Association, 19*(1), 6-12.

Li, S. Y., Tang, T.-L., Hickling, A., Yau, S., Brecknell, B., & Sanderson, P. M. (2017). Spearcons for Patient Monitoring: Laboratory Investigation Comparing Earcons and Spearcons. *Human Factors*, 0018720817697536.

Liu, D., Grundgeiger, T., Sanderson, P. M., Jenkins, S. A., & Leane, T. A. (2009). Interruptions and blood transfusion checks: lessons from the simulated operating room. *Anesthesia and Analgesia, 108*(1), 219-222. doi:10.1213/ane.0b013e31818e841a

Liu, D., Jenkins, S. A., & Sanderson, P. M. (2009). *Clinical implementation of a head-mounted display of patient vital signs.* Paper presented at the 2009 International Symposium on Wearable Computers, Proceedings.

Liu, D., Jenkins, S. A., Sanderson, P. M., Fabian, P., & Russell, W. J. (2010). Monitoring with head-mounted displays in general anesthesia: a clinical evaluation in the operating room. *Anesthesia and Analgesia, 110*(4), 1032-1038. doi:10.1213/ANE.0b013e3181d3e647

Liu, D., Jenkins, S. A., Sanderson, P. M., Watson, M. O., Leane, T., Kruys, A., & Russell, W. J. (2009). Monitoring with head-mounted displays: performance and safety in a full-scale simulator and part-task trainer. *Anesthesia and Analgesia, 109*(4), 1135-1146. doi:10.1213/ANE.0b013e3181b5a200

Loeb, R. G., & Fitch, W. T. (2002). A laboratory evaluation of an auditory display designed to enhance intraoperative monitoring. *Anesthesia and Analgesia, 94*(2), 362-368.

Monroy, G. L., Shemonski, N. D., Shelton, R. L., Nolan, R. M., & Boppart, S. A. (2014). *Implementation and evaluation of Google Glass for visualizing real-time image and patient data in the primary care office*.

Muensterer, O. J., Lacher, M., Zoeller, C., Bronstein, M., & Kübler, J. (2014). Google Glass in pediatric surgery: An exploratory study. *International Journal of Surgery, 12*(4), 281-289.

Mumaw, R. J., Roth, E. M., Vicente, K. J., & Burns, C. M. (2000). There is more to monitoring a nuclear power plant than meets the eye. *Human Factors, 42*(1), 36-55.

Mustonen, T., Berg, M., Kaistinen, J., Kawai, T., & Hakkinen, J. (2013). Visual task performance using a monocular see-through head-mounted display (HMD) while walking. *Journal of Experimental Psychology: Applied, 19*(4), 333-344. doi:10.1037/a0034635

Nikolic, M. I., Orr, J. M., & Sarter, N. B. (2004). Why pilots miss the green box: How display context undermines attention capture. *International Journal of Aviation Psychology, 14*(1), 39-52. doi:DOI 10.1207/s15327108ijap1401_3

Nikolic, M. I., & Sarter, N. B. (2001). Peripheral visual feedback: a powerful means of supporting effective attention allocation in event-driven, data-rich environments. *Human Factors, 43*(1), 30-38. doi:10.1518/001872001775992525

Ormerod, D., Ross, B., & Naluai-Cecchini, A. (2002). *Use of a see-through head-worn display of patient monitoring data to enhance anesthesiologists' response to abnormal clinical events.* Paper presented at the Wearable Computers, 2002.(ISWC 2002). Proceedings. Sixth International Symposium on.

Paine, C. W., Goel, V. V., Ely, E., Stave, C. D., Stemler, S., Zander, M., & Bonafide, C. P. (2016). Systematic Review of Physiologic Monitor Alarm Characteristics and Pragmatic

Interventions to Reduce Alarm Frequency. *Journal of Hospital Medicine, 11*(2), 136-144. doi:10.1002/jhm.2520

Pascale, M., Sanderson, P., Liu, D., Mohamed, I., & Loeb, R. (2015). *Event detection using a simulated head-worn display.* Paper presented at the Proceedings 19th Triennial Congress of the IEA.

Pascale, M., Sanderson, P., Liu, D., Mohamed, I., Stigter, N., & Loeb, R. (2015). *Peripheral detection for abrupt onset stimuli presented via head-worn display.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Paterson, E., Sanderson, P., Paterson, N., & Loeb, R. (2017). Effectiveness of enhanced pulse oximetry sonifications for conveying oxygen saturation ranges: a laboratory comparison of five auditory displays. *BJA: British Journal of Anaesthesia*, aex343.

Patterson, R., Winterbottom, M. D., & Pierce, B. J. (2006). Perceptual issues in the use of head-mounted visual displays. *Human Factors, 48*(3), 555-573. doi:Doi 10.1518/001872006778606877

Patterson, R., Winterbottom, M. D., Pierce, B. J., & Fox, R. (2007). Binocular rivalry and head-worn displays. *Human Factors, 49*(6), 1083-1096. doi:10.1518/001872007x249947

Rivera-Rodriguez, A., & Karsh, B.-T. (2010). Interruptions and distractions in healthcare: review and reappraisal. *BMJ Quality & Safety, 19*(4), 304-312.

Ruskin, K. J., & Hueske-Kraus, D. (2015). Alarm fatigue: impacts on patient safety. *Current Opinion in Anesthesiology, 28*(6), 685-690.

Sanderson, P. M. (2006). The multimodal world of medical monitoring displays. *Applied Ergonomics, 37*(4), 501-512. doi:10.1016/j.apergo.2006.04.022

Sasangohar, F., Donmez, B., Trbovich, P., & Easty, A. C. (2012). *Not all interruptions are created equal: positive interruptions in healthcare.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Sawyer, B. D., Finomore, V. S., Calvo, A. A., & Hancock, P. A. (2014). Google glass: A driver distraction cause or cure? *Human Factors, 56*(7), 1307-1321. doi:10.1177/0018720814555723

Schlosser, P., Sanderson, P., Grundgeiger, T., Liu, D., & Loeb, R. G. (2016). *The Effect of Conventional Screens vs. Head-Mounted Displays on Alarm Monitoring Strategies.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Seagull, F. J., Wickens, C. D., & Loeb, R. G. (2001). *When is less more? Attention and workload in auditory, visual, and redundant patient-monitoring conditions.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Senders, J. W. (1964). The human operator as a monitor and controller of multidegree of freedom systems. *Human Factors in Electronics, IEEE Transactions on*(1), 2-5.

Sheridan, T. B. (1970). On how often the supervisor should sample. *IEEE Transactions on Systems Science and Cybernetics, 6*(2), 140-145.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception, 28*(9), 1059-1074. doi:Doi 10.1068/P2952

Sitterding, M. C., Broome, M. E., Everett, L. Q., & Ebright, P. (2012). Understanding situation awareness in nursing work: a hybrid concept analysis. *Advances in Nursing Science, 35*(1), 77-92.

Solet, J. M., & Barach, P. R. (2012). Managing alarm fatigue in cardiac care. *Progress in Pediatric Cardiology, 33*(1), 85-90. doi:10.1016/j.ppedcard.2011.12.014

Sowan, A. K., Gomez, T. M., Tarriela, A. F., Reed, C. C., & Paper, B. M. (2016). Changes in Default Alarm Settings and Standard In-Service are Insufficient to Improve Alarm Fatigue in an Intensive Care Unit: A Pilot Project. *JMIR Hum Factors, 3*(1), e1. doi:10.2196/humanfactors.5098

Sowan, A. K., & Reed, C. C. (2017). A Complex Phenomenon in Complex Adaptive Health Care Systems-Alarm Fatigue. *JAMA Pediatr, 171*(6), 515-516. doi:10.1001/jamapediatrics.2016.5137

Sowan, A. K., Tarriela, A. F., Gomez, T. M., Reed, C. C., & Rapp, K. M. (2015). Nurses' Perceptions and Practices Toward Clinical Alarms in a Transplant Cardiac Intensive Care Unit: Exploring Key Issues Leading to Alarm Fatigue. *JMIR Hum Factors, 2*(1), e3. doi:10.2196/humanfactors.4196

Stubbings, L., Chaboyer, W., & McMurray, A. (2012). Nurses' use of situation awareness in decision-making: an integrative review. *Journal of Advanced Nursing, 68*(7), 1443-1453.

Taenzer, A. H., Pyke, J. B., & McGrath, S. P. (2011). A review of current and emerging approaches to address failure-to-rescue. *Anesthesiology, 115*(2), 421-431. doi:10.1097/ALN.0b013e318219d633

Tang, T.-I. (2016). *Can I Get a Heads-Up On That Patient? Head-Worn Displays For Patient Monitoring.* (BS (Hons)), The University of Queensland, St Lucia, Australia.

Theeuwes, J., Olivers, C. N., & Belopolsky, A. (2010). Stimulus-driven capture and contingent capture. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(6), 872-881.

Tippey, K. G., Sivaraj, E., Ardoin, W.-J., Roady, T., & Ferris, T. K. (2014). *Texting while driving using Google Glass Investigating the combined effect of heads-up display and hands-free input on driving safety and performance.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Vorraber, W., Voessner, S., Stark, G., Neubacher, D., DeMello, S., & Bair, A. (2014). Medical applications of near-eye display devices: an exploratory study. *International Journal of Surgery (London, England), 12*(12), 1266-1272. doi:10.1016/j.ijsu.2014.09.014

Watkins, T., Whisman, L., & Booker, P. (2016). Nursing assessment of continuous vital sign surveillance to improve patient safety on the medical/surgical unit. *Journal of Clinical Nursing, 25*(1-2), 278-281. doi:10.1111/jocn.13102

Watson, M., & Sanderson, P. (2004). Sonification supports eyes-free respiratory monitoring and task time-sharing. *Human Factors, 46*(3), 497-517.

Wickens, C. D., Dixon, S. R., & Seppelt, B. (2005). *Auditory preemption versus multiple resources: Who wins in interruption management?* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

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. doi:DOI 10.1518/hfes.45.3.360.27250

Wickens, C. D., Rice, S., Keller, D., Hutchins, S., Hughes, J., & Clayton, K. (2009). False alerts in air traffic control conflict alerting system: Is there a “cry wolf” effect? *Human Factors, 51*(4), 446-462.

Wilken, M., Hüske-Kraus, D., Klausen, A., Koch, C., Schlauch, W., & Röhrig, R. (2017). Alarm Fatigue: Causes and Effects. *Studies in Health Technology and Informatics, 243*, 107.

Winterbottom, M. D., Patterson, R., Pierce, B. J., Covas, C. M., & Rogers, J. (2006). *Binocular rivalry and attention in helmet-mounted display applications.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Winterbottom, M. D., Patterson, R., Pierce, B. J., Covas, C. M., & Winner, J. (2007). Depth of focus and visual recognition of imagery presented on simultaneously viewed displays: implications for head-mounted displays. *Human Factors, 49*(5), 907-919. doi:10.1518/001872007x230253

Winterbottom, M. D., Patterson, R., Pierce, B. J., Gaska, J., & Hadley, S. (2015). *Visibility of monocular symbology in transparent head-mounted display applications.* Paper presented at the SPIE Defense+ Security.

Winterbottom, M. D., Patterson, R., Pierce, B. J., & Taylor, A. (2006). *Visual suppression of monocularly presented symbology against a fused background in a simulation and training environment.* Paper presented at the Defense and Security Symposium.

Wolfe, J. M., O’Neill, P., & Bennett, S. C. (1998). Why are there eccentricity effects in visual search? Visual and attentional hypotheses. *Perception and Psychophysics, 60*(1), 140-156.

Woodham, A., Billinghurst, M., & Helton, W. S. (2016). Climbing With a Head-Mounted Display: Dual-Task Costs. *Human Factors, 58*(3), 452-461. doi:10.1177/0018720815623431

Woods, D. D. (1995). The Alarm Problem and Directed Attention in Dynamic Fault Management. *Ergonomics, 38*(11), 2371-2393. doi:Doi 10.1080/00140139508925274

Xiao, Y., Seagull, F. J., Nieves-Khouw, F., Barczak, N., & Perkins, S. (2004). Organizational-historical analysis of the" failure to respond to alarm" problems. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 34*(6), 772-778.

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance, 10*(5), 601-621.

Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance, 16*(1), 121-134. doi:10.1037//0096-1523.16.1.121

Zhang, Y., Drews, F., Westenskow, D. R., Foresti, S., Agutter, J., Bermudez, J. C., . . . Loeb, R. (2002). Effects of integrated graphical displays on situation awareness in anaesthesiology. *Cognition, Technology & Work, 4*(2), 82-90.

Appendices

Appendix A: Manuscript 1

Pascale, M., Sanderson, P., Liu, D., Mohamed, I., Stigter, N., Loeb, R. (under review). Detection of visual stimuli on monocular peripheral head-worn displays. *Applied Ergonomics. Submitted 13 December 2017.*

Detection of visual stimuli on monocular peripheral head-worn displays

Michael T. Pascale[1]
Penelope Sanderson[1 2 3]
David Liu[2 3]
Ismail Mohamed[1 2]
Nicola Stigter[1]
Robert G. Loeb[4]

[1]School of Psychology, The University of Queensland, Queensland, Australia
[2]School of ITEE, The University of Queensland, Queensland, Australia
[3]School of Medicine, The University of Queensland, Queensland, Australia
[4]College of Medicine, University of Florida, Gainesville, Florida, USA

Running head: Peripheral detection performance with head-worn displays
Manuscript type: STUDY
Abstract: 248
Word Count: 4712
Declarations of interest: none

Corresponding author: Michael T. Pascale – michael.pascale@uq.net.au
Present Address: School of Psychology, Sir Fred Schonell Dr., St Lucia, QLD 4072, AUS

Abstract

Objective: To compare people's ability to detect peripherally presented stimuli on a monocular head-worn display (HWD) versus a conventional screen.

Background: Visual attention capture has been systematically investigated, but not with respect to HWDs. How stimulus properties affect attention capture is likely to be different on an HWD when compared to a computer display.

Method: Participants performed an ongoing perceptual task and attempted to detect stimuli that were displayed peripherally on either a computer monitor or a monocular HWD.

Results: Participants were less able to detect peripheral stimuli when the stimuli were presented on an HWD than when presented on a computer monitor. Moreover, the disadvantage of the HWD was more pronounced when peripheral stimuli were less distinct and when the stimuli were presented further into the periphery

Conclusion: Presenting stimuli on a monocular head-worn display reduces participants' ability to notice peripheral visual stimuli compared to presentation on a normal computer monitor. This effect increases as stimuli are presented further in the periphery, but can be ameliorated to a degree by using high-contrast stimuli.

Application: The findings are useful for designers creating visual stimuli intended to capture attention when viewed on a peripherally positioned monocular head-worn display.

Keywords: Head-worn displays, Google Glass, monitoring, attention, perception.

Précis: Several factors can change the noticeability of stimuli when viewed peripherally: distance of stimuli from the primary task, stimulus brightness and degree of tilt. These effects are stronger when peripheral stimuli are displayed on a monocular HWD than on a screen.

Introduction

When people are engaged in mobile work, head worn displays (HWDs) can provide real-time access to information that might otherwise be unavailable or difficult to access. For example, an anesthetist can continually monitor the HWD for changes in a patient's vital signs, rather than visually scanning equipment around the room. Alternatively, the HWD could alert the user to something important happening away from the current task. For example, a nurse focusing on medication preparation in one location may be notified, via HWD, that a patient in another location has a critically low heart rate. The nurse may not be actively attending to the HWD, but their attention to the HWD may be triggered by the change. In that case, it would be important to know how visual stimuli on an HWD capture attention to ensure recognition of the change. The purpose of the experiment reported in this paper was to compare people's ability to detect peripherally-located stimulus changes across two display mediums, specifically a monocular see-through HWD versus a conventional computer screen.

A distinction has been made between goal-driven attention (voluntary or endogenous) and stimulus driven attention (involuntary or exogenous) (Folk, Remington, & Johnston, 1992). On one hand, exogenous attention capture can be driven by stimulus related factors. For example, color, brightness, or motion may be manipulated to enhance visual discriminability, thereby increasing the likelihood of capturing visual attention (Hillstrom & Yantis, 1994; Nikolic, Orr, & Sarter, 2004; Yantis & Jonides, 1984). Likewise, increases in the peripheral eccentricity (distance from foveal vision) of the stimuli will reduce people's ability to notice target changes due to the organization of receptors on the retina (Nikolic et al., 2004; Olzak & Thomas, 1986; Wolfe, O'Neill, & Bennett, 1998). On the other hand, factors affecting exogenous attention capture can be task related. For example, perceptual tasks will reduce people's ability to detect task-irrelevant stimuli more so than cognitive tasks (Lavie, 2005, 2010; Lavie, Beck, & Konstantinou, 2014). Taken together, the above studies provide a basis for designing HWD stimuli that can effectively capture attention, but if characteristics of a head-worn device introduce additional limitations, special consideration may be needed.

To date, only a few studies have investigated attention capture with HWDs. Using a foveal HWD, Winterbottom, Patterson, Pierce, Gaska, and Hadley (2015) found that target stimuli presented directly in the forward field of view were detected less often when they were presented via monocular HWD, compared to binocular; and that the stimuli required greater visual contrast thresholds to attract attention. Reductions in detection rates were found in the extreme periphery as well, but using a different kind of head-worn device. Costanza, Inverso, Pavlov, Allen, and Maes (2006) showed that increases in task loading reduced how effectively

attention was captured by an array of light emitting diodes (LEDs) located peripherally, at the hinge of a normal pair of regular glasses. The relationship between HWD stimuli and peripheral eccentricity, however, is still unknown, especially when compared to detection rates for peripheral stimuli displayed on a computer display.

Regardless of their location on the visual field, stronger visual stimuli (more intensity), or the addition of an auditory cure, can be used to capture attention more reliably. Woodham, Billinghurst, and Helton (2016) found that rock climbers were less likely to notice words presented on a monocular, peripheral HWD (Google Glass, Google Inc., Mountain View, CA, USA) while climbing (versus sitting) unless they were presented with a simultaneous auditory cue. This is because auditory cues have a pre-emptive quality (Wickens, 2008; Wickens, Dixon, & Seppelt, 2005) that can aid target detection, but for that reason, they can also be potentially distracting, and are not always appropriate or desirable, particularly in an environment that is already rich with sounds, like a hospital ward.

Moreover, HWDs that attract too much attention may compromise participants' ability to perform their ongoing task, just as wearing one HWD did in simulated driving tasks (Chua, Perrault, Matthies, & Zhao, 2016; Sawyer, Finomore, Calvo, & Hancock, 2014). Furthermore, Mustonen, Berg, Kaistinen, Kawai, and Hakkinen (2013) found that participants attempting to detect changes on an HWD were significantly affected by whether or not they were seated or walking. Specifically, the dual-task requirements of walking and attempting to view the HWD resulted in more missed targets, but also more walking errors.

The current study aimed to extend the questions answered above to simple changes to target stimuli presented on a peripheral, monocular HWD. Specifically, the goal was to compare participants ability to detect visual changes on the HWD to their ability to detect equivalent changes on a conventional computer screen. The study we report was designed to examine participants' performance when peripheral stimuli of differing brightness and orientation were presented on a simulated HWD (computer screen, binocular) versus on a real HWD (monocular), and at near versus far eccentricities. We predicted that the probability of detecting target stimuli would be significantly reduced (a) when participants viewed the peripheral stimuli on the real HWD rather than on the simulated HWD, (b) when the peripheral stimuli were at the far eccentricity rather than near, as in Nikolic et al. (2004) and Wolfe et al. (1998), and (c) when target stimuli shared more characteristics with the default, vertical, non-target stimuli, following Jonides (1981), Yantis and Jonides (1984), and Hillstrom and Yantis (1994).

Method

Participants

72 students from The University of Queensland participated in exchange for $10 gift cards. The sample size was determined by a power analysis using the results of a pilot study with a comparable design (M = .642, *Mdelta* = .558, SD = .362, r = .74, α = .05, β = .80). Applicants wearing corrective eyeglasses were excluded prior to enrollment. This research complied with the American Psychological Association Code of Ethics and was approved by the Institutional Review Board at The University of Queensland. Informed consent was obtained from each participant.

Design

The current experiment was within-subjects exploring the effects on peripheral target detection of display medium (simulated HWD versus real HWD), peripheral eccentricity (near versus far), apparent motion of target stimuli (none versus slight change from vertical to tilted), and brightness of target stimuli (dark gray versus light gray versus white). The study was conducted in eight blocks of trials. During each block, participants performed eight six-minute trials of an ongoing task presented on a computer monitor while they detected changes to peripheral stimuli on either the computer monitor (simulated HWD) or the real HWD. The changes occurred at random-appearing intervals.

Apparatus

The participant sat in an adjustable chair in front of a computer monitor, which was positioned on a small stand (10 cm high). The participant maintained a constant viewing distance from the center of the computer monitor by resting their chin in a chinrest. The distance of the chinrest to the monitor screen remained constant (51 cm) across participants. The participant's head was further stabilized using a headrest. They positioned their forehead against the headrest to maintain the angle at which the image on the HWD would be seen against the background of the computer, when the HWD was worn. Both the chinrest and headrest were adjusted vertically for each participant so that the HWD image overlaid, as closely as possible, the position on the screen where the simulated HWD was otherwise presented.

The ongoing task and simulated HWD stimuli were presented on a 27-inch iMac computer display with a black background and a calibrated background (see below), respectively. The real HWD stimuli were presented on Google Glass (Google Inc., Mountain

View, CA), a monocular see-through head-worn display that is visible to only the right eye, but not directly in the forward field of view.

The participant made their inputs on a standard keyboard whose relevant keys were covered with a small patch of Velcro, fuzzy side up, to make them easier for the participant to feel and therefore use without having to look down to check their finger location.

Calibration

Before the experiment was conducted, a calibration study was run to equalize, as much as possible, the colors and brightness of the simulated HWD display on the computer screen and of the real HWD display. A black background on the HWD does not actually look black, but instead appears as a desaturated rust color (approximated in Figure 1) that was also affected by whatever it overlaid in the environment—in our case, the black screen of the computer monitor. In addition, before the calibration, the stimuli on the HWD appeared less bright than those on the computer display.

A sample of participants separate from those in the main study viewed both displays and made color adjustments to the background of the simulated HWD displayed on the computer monitor. Their goal was to match the background color of the simulated HWD to the background color of the real HWD. The participant then adjusted the brightness of the stimuli on the real HWD to match their brightness on the simulated HWD. Because the distance between the HWD display and the computer screen was fixed, given the chinrest and headrest, the stimuli on the simulated HWD and real HWD appeared to be the same size and at the same degree of eccentricity from forward gaze. The calibration study produced color values and intensities for use in the current experiment that equalized the colors and brightness of the stimuli on the two displays. The procedure and full findings of the calibration study are reported in Appendix A.

Stimuli and Tasks

For all blocks of the current study, participants performed an ongoing perceptual task presented in yellow font in one of two locations on the computer monitor (near or far) (see Figure 1). Two squares (Landolt stimuli) were displayed side-by-side. Each square had a disconnected side with gaps of different widths. Using their left hand, participants indicated whether the gap in the left or right square was larger, by pressing the “A” key for the left square, or the “S” key for the right square. Participants were instructed to focus on this task. The task was self-paced but stimuli advanced if no response was received within 6 seconds.

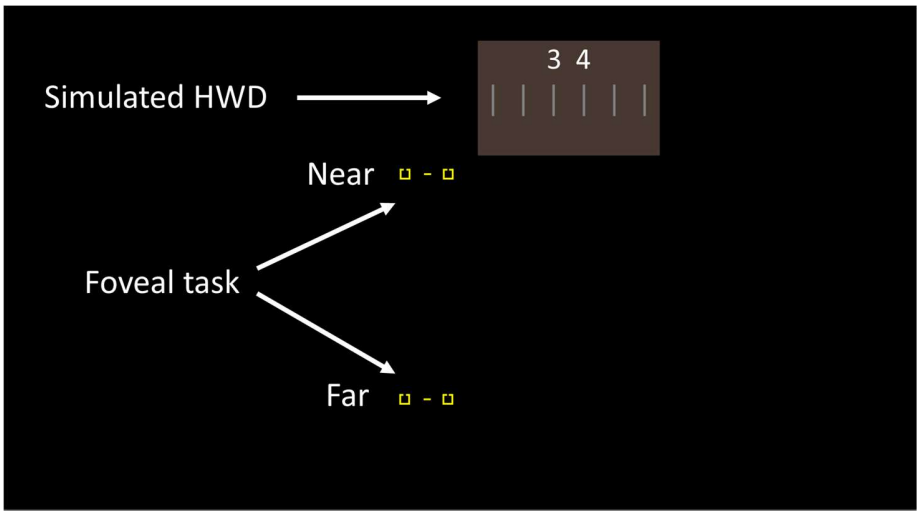

Figure 1. Simulated HWD stimuli appeared in a constant screen location, and were designed to be viewed in peripheral vision. One of the middle bars (3 or 4) would briefly change to one of the five target states in Figure 2. The ongoing task was presented at either a near or far location with respect to the peripheral stimuli. Annotations, arrows, and bar numbers 3 and 4 were not shown during the experiment.

Table 1

Measure of the visual angles from the foveal task to each bar on the peripheral display, at each eccentricity

Eccentricity	Visual Angle (degrees)					
	Bar 1	Bar 2	Bar 3	Bar 4	Bar 5	Bar 6
NEAR task	9	10.2	11.8	13.5	15.2	17.1
FAR task	24.3	24.7	25.3	26	26.9	27.8

The peripheral detection task was in a constant location above and to the right of the ongoing task, presented on either the simulated HWD or real HWD (Figure 1; visual angles from the foveal task to each bar are listed in Table 1). The stimuli to be detected in the periphery comprised six vertical bars that simulated a simple multiple-process display that might be seen on an HWD. In their default state, the bars were vertical and dark gray. Occasionally, a target stimulus was generated by briefly changing the tilt and/or brightness of the third or fourth bar. The target stimulus was an uppercase “I” in Lucida Grande font, either roman or italicized with a tilt of 11.5 degrees from vertical to produce “*I*”. Brightness levels on the simulated HWD were dark gray (hex code: #808080), light gray (hex code: #c0c0c0), or white (hex code: #ffffff). Altogether, there were five kinds of target stimuli, varying in brightness and/or tilt from the dark

gray and vertical default bars: dark gray and tilted, light gray and vertical, light gray and tilted, white and vertical, and white and tilted (see Figure 2).

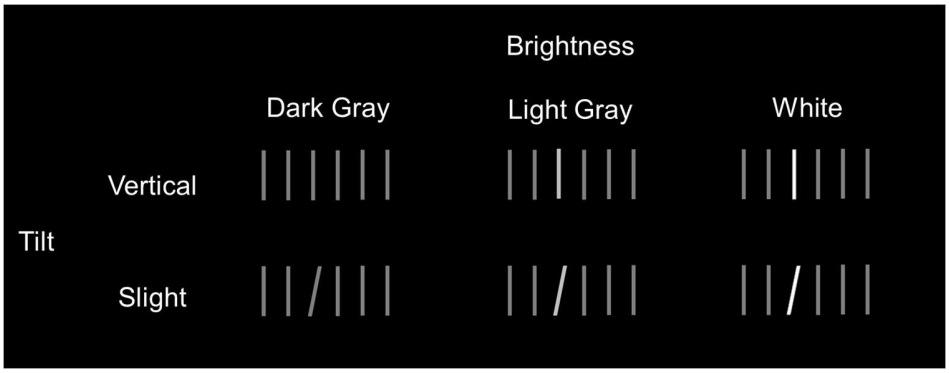

Figure 2. Examples of the five possible target changes in Experiment 1 compared with the default, non-target state (top left).

In each six-minute experimental block there were 15 appearances of target stimuli (three of each kind of stimulus) that were un-cued and that occurred between 17.1 and 30.9 seconds apart, with a mean of approximately 24 seconds apart. Target stimuli were timed to appear 400 milliseconds after the onset of an ongoing task trial and they persisted for 200 milliseconds before returning to the default state. The duration of 200ms was chosen because it is faster than the average time to react to and make a saccade from the ongoing task to the HWD, allowing us to measure the detection of a change when the HWD display was viewed in peripheral vision, and not when viewed directly. Using their right hand, participants pressed the semicolon key when they detected a target stimulus.

Procedure

The HWD was offered to the participant who put it on. In the real HWD condition the participant adjusted the angle of the HWD display screen until the stimuli were clearly visible. Participants then adjusted the height of their chair and location of the headrest to overlay the real HWD display comfortably over the simulated HWD so that the location of the display, relative to the foveal task, was matched across the two display conditions at test. Participants were instructed to maintain that position throughout each block.

Instructions and practice trials were presented using a timed Microsoft PowerPoint presentation with a recorded narration. Participants then worked through the eight experimental blocks, in which they performed both the ongoing task and peripheral target change detection task with either the simulated or the real HWD. Before the start of each block, the software instructed the participant to put on or take off the HWD in preparation for the upcoming block. Participants

were given one-minute breaks between blocks. The entire experiment lasted approximately 60 minutes.

Analyses

Analyses were directed at identifying whether participants' ability to detect targets (changes in peripheral stimuli) was affected by the display mode, the eccentricity of the HWD stimuli from the foveal task, and the deviation (in brightness and/or tilt) of the target from the dark gray and vertical default. Initially, the detection data (target detected or target missed) were transformed to accuracy values (percentage detected) for each combination of the independent variables. An ANOVA was used to analyze these data, but subsequent analyses of the residuals revealed that the data did not fit a normal distribution and did not have equal variances. Thus, target detection data were analyzed using a mixed effects logistic regression, which does not rely on normally distributed or homoscedastic data, and, furthermore, is the method best suited to handle the original binary outcome variable.

The final logistic regression model was selected based on the best fit to the data, as described in Marewski and Olsson (2009). Thus, in the model to be presented, any omitted interactions did not significantly influence the explanatory power of the model, but may have influenced the strength of the slope-coefficient of the remaining variables. The fixed effects of the final model included display, eccentricity, target stimuli, and the interactions (denoted by a "x" in the table) between display/eccentricity, and display/stimulus, with participants as a random effect. The regression model was then used to generate the predicted probability that each color/tilt combination would be detected on each display, and at each eccentricity.

Results

Table 2 shows the results of the logistic regression. Figure 3 shows the predicted probabilities of detecting each target stimulus when displayed on the simulated HWD versus the real HWD, and at the near versus far eccentricity. The baseline condition was coded as the dark gray and tilted target stimulus, on the simulated HWD, at the near eccentricity. Each line in Table 2 describes a specific deviation from that baseline set of conditions, and the subsequent change in the odds of detection (fixed effects). The last row reports the inter-participant variability (random effect).

The data show that the odds of detecting a target stimulus on the real HWD were significantly less than for the simulated HWD, for each kind of target stimulus. Moreover, when target stimuli were presented further into the periphery on the HWD, detection rates were reduced

to a greater degree than when presented further into the periphery on the simulated HWD. However, as stimuli increased in brightness and tilt, the likelihood of being detecting increased.

Table 2

Results of the mixed effects logistic regression, "x" indicates interactions

Change from baseline (simulated HWD, near, straight, and dark gray)	Odds Ratio	SE	z	p
Intercept (baseline)	6.44	1.06	11.36	< .001
Real HWD	0.11	0.02	-14.28	< .001
Far	0.64	0.08	-3.84	< .001
Real HWD x Far	0.51	0.08	-4.38	< .001
Light gray and vertical	1.36	0.20	2.11	0.035
Light gray and tilted	3.62	0.65	7.13	< .001
White and vertical	3.85	0.71	7.33	< .001
White and tilted	10.16	2.56	9.20	< .001
Real HWD x Light gray and vertical	7.27	1.43	10.07	< .001
Real HWD x Light gray and tilted	4.34	0.99	6.44	< .001
Real HWD x White and vertical	5.04	1.18	6.92	< .001
Real HWD x White and tilted	3.31	0.99	4.01	< .001
Random effect - Participant	0.75	0.16		

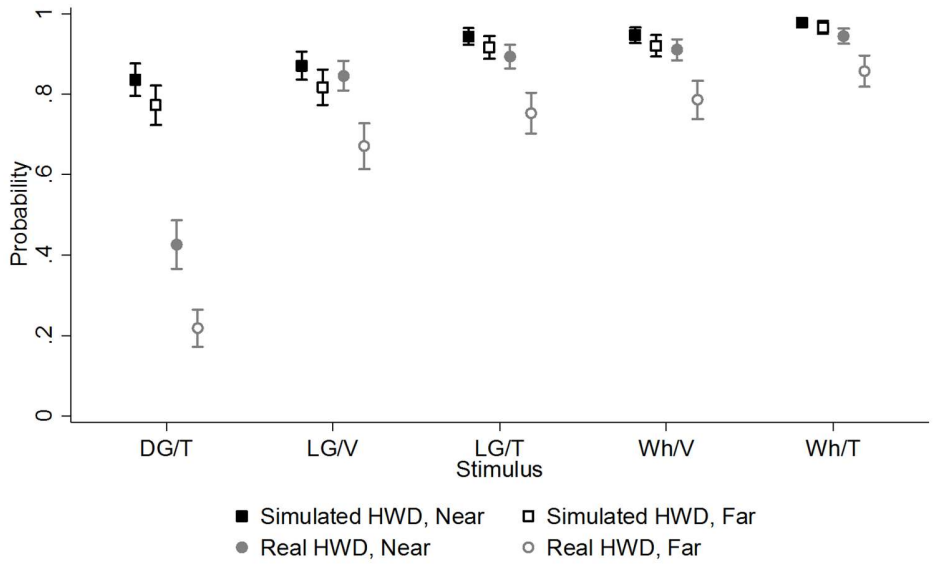

Figure 3. Predicted probability and 95% confidence intervals of detecting each target stimulus on the simulated HWD versus the real HWD, and at the near versus far eccentricity. DG/T = dark gray and tilted, LG/V = light gray and vertical, LG/T = light gray and tilted, Wh/V = white and vertical, and Wh/T = white and tilted.

Discussion

The purpose of this study was to investigate differences in people's ability to detect changes to peripheral stimuli displayed on a real HWD versus their ability to detect matched stimuli on a regular computer screen simulating an HWD. The size of targets, contrast from background, and position in the visual field were all controlled to be as similar as possible across the real and simulated HWD displays. By equating colors and brightness across displays as much as possible, our goal was to lessen the chance that any difference in detection rates between the simulated HWD and real HWD conditions would be due to simple device-related differences such as target size, brightness, color, or color contrasts between the stimuli and the background. Accordingly, any difference in detection rates would probably reflect differences unique to the two display devices themselves.

Our hypotheses were confirmed. The HWD reduced participants' ability to detect peripheral targets when compared with detection using the simulated HWD display. Moreover, the disadvantage of the HWD was stronger when targets were farther in peripheral vision and when targets were only minimally different from their default dark gray and vertically oriented form. However, the disadvantage of the HWD was nearly eliminated when targets were at their most distinct, such as when they were light gray and tilted, and when they were white and tilted. There are several possible explanations for the findings that are suggested by prior research in the area. One explanation for the differences in detection performance between the simulated HWD and the real HWD may lie in the fact that the real HWD stimuli were seen by only one eye. When stimuli are seen by both eyes simultaneously, the images are fused resulting in binocular summation (Blake & Fox, 1973), which results in gains in contrast sensitivity, brightness sensitivity, flicker perception and visual acuity. Stimuli presented to only one eye are subject to binocular rivalry (Patterson, Winterbottom, Pierce, & Fox, 2007)—the unconscious prioritization, and continual switching, of stimulus perception from one eye over the other. If the image from the eye viewing the real HWD had been temporarily suppressed when the target stimuli were being displayed, it would account for the reduction in target detection rates on that display.

A second explanation for the differences in detection between the simulated vs. real HWD may be that distant fixated objects take perceptual priority over proximate obstructions (Arnold (2011). This explanation is similar to that for binocular rivalry, but proposes that suppression is related to signal strength, and depends on whether or not the images from each eye match. For our study, the real HWD obstructed part of the view of the computer screen that displayed the perceptual task, but only on the image from the right eye. As a result, the visual system may have intentionally suppressed that area of the visual field of the right eye (the

location of the real HWD), to favor the view of the fixated screen and task, reducing the likelihood that stimuli on the real HWD would be detected.

A third explanation for the differences between detection rates on the simulated vs. real HWD may lie in the differences between the focal plane of the target stimuli on the real HWD and the focal plane of the ongoing task. Winterbottom, Patterson, Pierce, Covas, and Winner (2007) found that focal depth can affect the detectability of stimuli as well as visual comfort, and that the optimal focal depth should be the midpoint of the range of distances of potential stimuli. These distinctions, however, are for HWD imagery presented directly in the forward field of view, rather than peripherally.

A fourth explanation for the difference in detection rates between simulated vs. real HWDs might be that the relationship between the ongoing task and peripheral stimuli was always fixed when participants viewed the peripheral stimuli on the simulated HWD, but variable when viewed on the real HWD. When participants viewed the peripheral stimuli on the real HWD, head motion was constrained by the chinrest and headrest, but the stimuli could still move slightly with respect to the ongoing task when small head movements occurred, which may have reduced participants' ability to perceive the targets. According to Gestalt theory, stimuli that have common fate or 'shared movement', are chunked together and experienced preattentively as being a single object (Koffka, 1935). Attention, chunking can influence how efficiently our visual system attends to objects in view, and whether or not attention needs to be divided across multiple perceptual objects (Duncan, 1984). This might suggest that the simulated HWD and ongoing task were viewed as a single object because they were viewed on the same computer screen, thereby not requiring a division of attention. The imagery on the real HWD, however, may have functioned as a separate object from the ongoing task, causing a greater division in attention load than for the simulated HWD. Therefore, the items on the real HWD may not have been monitored as effectively as the items on the simulated HWD.
The differences in the rates of detection between stimuli were driven by the differences in contrast. Stimuli that share fewer similarities with the background, stand out to the perceptual system to a greater degree, for example, a black bear walking on snow versus a polar bear walking on snow. Our data extend previous findings of the effects of contrast on computer displays (Yantis & Jonides, 1984) and HWDs (Winterbottom et al., 2015) by revealing an increased risk of detection failure for stimuli that are very low in contrast. The dark gray and tilted stimulus (the stimulus with the least amount of contrast) in the current experiment was detected least often, but the effect was much stronger on the HWD than with the computer display.

An explanation for the differences between near vs. far stimuli lies in the fact that peripheral stimuli naturally receive less attention than stimuli positioned at fixation. This is because there are fewer physical receptors on peripheral areas of the retina, but also because there is an attentional bias for stimuli in central areas of the retina (Wolfe et al., 1998).

Putting it all together, the hyperadditive disadvantage seen for peripheral stimuli on the real HWD may be explained by an increase in peripheral suppression for monocularly viewed stimuli. Blake, O'Shea, and Mueller (1992) have reported that peripheral zones of rivalry tend to be larger than foveal zones. When combined with the perceptual priority phenomenon described in Arnold (2011), the binocularly unmatched image from the real HWD in the right eye would be expected to be more susceptible to suppression when it is viewed further into the periphery. Stimuli high in contrast from the background can break through some of that suppression, but as contrast from the background decreases, the stimulus is more susceptible to the increased suppression.

Whatever the reasons for worsened performance with the real HWD, our findings show that peripheral, monocular, optical see-through HWDs may require stronger stimuli to attract exogenous attention than are needed with a conventional screen, especially when the HWD is further in the periphery. Despite this concern, it may still be better for mobile workers to use an HWD than not to do so. Without an HWD, a mobile, multitasking worker might have no information at all about a process they must monitor, or the information might not be ready to hand, which is arguably worse than having access to information on an HWD that is sometimes indistinct and therefore occasionally missing a cue. In a nursing context, for example, with an HWD, clinically-relevant patient changes could be signaled in a way that ensures they would be noticed, but without the kind of auditory alerts that could contribute to alarm fatigue (Cvach, 2012; Graham & Cvach, 2010; Ruskin & Hueske-Kraus, 2015), and without the distracting effects commonly see on ongoing task that performance while using HWDs (Chua et al., 2016; Sawyer et al., 2014; Theis, Mertens, & Wille, 2015).

Limitations.

One limitation of the current experiment is that it tested participants' ability to detect simple visual changes occurring over a very short period. The stimuli we tested are not intended literally as designs for representing changes that might happen within a monitored system. For example, changes in a work context might occur as a trend over seconds or minutes. Prolonged changes, however, can still be missed. For example, Liu, Jenkins, and Sanderson (2009) found that anesthetists failed to recognize a slow change in a waveform depicting a change to a patient's condition, which the authors attributed in part to inattentional blindness (Simons &

Chabris, 1999). Strong changes in brightness or apparent motion may still be an effective way to attract attention to such trends, but further work is needed to explore different display configurations and stimuli.

A further limitation is that the current study focused on identifying conditions for effective exogenous orienting of attention, but endogenous orienting of attention towards the HWD display could potentially develop after some use and familiarity with the behavior of the information sources. Many studies have shown that monitoring behavior tends to adapt to the likelihood of relevant information from a source (Sheridan, 1970). For patient monitoring, relevant information may be the relative likelihood that a patient's status will deteriorate. Thus, as familiarity with specific patients and their vital signs increases, clinicians may not need to rely on exogenous cueing. Instead they may notice deteriorations sooner, before a warning or critical threshold is met, and thus may be able to treat them sooner, and with more diagnostic context.

Future considerations

HWDs offer advantages and disadvantages for monitoring tasks. In many work environments, such as those where a worker is mobile or in a specialized space, a normal computer screen may not be continuously present or ready to hand. As a result, any amount of information that an HWD offers may be better than none. In that case, the disadvantages of an HWD compared with a conventional screen might seem to be a minor concern. However, if participants fail to notice certain stimuli on an HWD, while depending on it for information, the level of concern increases. Considerable further research is required to determine how users might use an HWD to monitor for trends rather than abrupt changes, and what the best interplay is of exogenous and endogenous cueing of attention via both visual and auditory channels. With such information, the designer will have a more theoretically grounded set of principles with which to design stimuli and displays for HWDs used in work contexts where mobile users must handle multiple simultaneous tasks.

Key Points

- Stimuli presented via monocular HWDs are generally less noticeable than when presented via a normal computer display.
- Stimuli with low visual contrast are significantly less noticeable on a monocular see-through HWD than on a normal computer display.
- Visual stimuli presented vian HWD, alone, may not be sufficient in capturing attention.
- Care is needed when designing displays to be monitored on head-worn devices to ensure appropriate capture of attention.

Acknowledgements

Funding for this project was provided by the Australian Research Council Discovery Project DP140101822 to Sanderson, Loeb, and Liu. We thank all the members of the Cognitive Engineering Research Group for their feedback on earlier versions of this manuscript. The experiment described in this manuscript was reported in abbreviated form in Pascale et al. (2015).

References

Arnold, D. (2011). I Agree: Binocular Rivalry Stimuli are Common but Rivalry is Not. *Frontiers in Human Neuroscience, 5*(157). doi:10.3389/fnhum.2011.00157

Blake, R., & Fox, R. (1973). The psychophysical inquiry into binocular summation. *Perception and Psychophysics, 14*(1), 161-185. doi:10.3758/bf03198631

Blake, R., O'Shea, R. P., & Mueller, T. J. (1992). Spatial zones of binocular rivalry in central and peripheral vision. *Visual Neuroscience, 8*(5), 469-478.

Chua, S. H., Perrault, S. T., Matthies, D. J., & Zhao, S. (2016). *Positioning glass: investigating display positions of monocular optical see-through head-mounted display*. Paper presented at the Chinese CHI, San Jose, CA, USA.

Costanza, E., Inverso, S. A., Pavlov, E., Allen, R., & Maes, P. (2006). *Eye-q: Eyeglass peripheral display for subtle intimate notifications.* Paper presented at the Proceedings of the 8th conference on Human-computer interaction with mobile devices and services.

Cvach, M. (2012). Monitor alarm fatigue: an integrative review. *Biomedical Instrumentation and Technology, 46*(4), 268-277. doi:10.2345/0899-8205-46.4.268

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General, 113*(4), 501-517. doi:10.1037/0096-3445.113.4.501

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology-Human Perception and Performance, 18*(4), 1030-1044. doi:Doi 10.1037//0096-1523.18.4.1030

Graham, K. C., & Cvach, M. (2010). Monitor alarm fatigue: Standardizing use of physiological monitors and decreasing nuisance alarms. *American Journal of Critical Care, 19*(1), 28-34. doi:10.4037/ajcc2010651

Hillstrom, A. P., & Yantis, S. (1994). Visual motion and attentional capture. *Attention, Perception, & Psychophysics, 55*(4), 399-411.

Jonides, J. (1981). Voluntary vs. automatic control over the mind's eye's movement. In J. B. Long & A. D. Baddelay (Eds.), *Attention and Performance IX* (pp. 187-203). Hillside, NJ: Erlbaum.

Koffka, K. (1935). *Principles of Gestalt psychology* (Vol. 44): Routledge.

Lavie, N. (2005). Distracted and confused?: selective attention under load. *Trends Cogn Sci, 9*(2), 75-82. doi:10.1016/j.tics.2004.12.004

Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current Directions in Psychological Science, 19*(3), 143-148. doi:Doi 10.1177/0963721410370295

Lavie, N., Beck, D. M., & Konstantinou, N. (2014). Blinded by the load: attention, awareness and the role of perceptual load. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 369*(1641), 20130205. doi:10.1098/rstb.2013.0205

Liu, D., Jenkins, S. A., & Sanderson, P. M. (2009). Patient monitoring with head-mounted displays. *Current Opinions in Anaesthesiology, 22*(6), 796-803. doi:10.1097/ACO.0b013e32833269c1

Marewski, J. N., & Olsson, H. (2009). Beyond the null ritual: Formal modeling of psychological processes. *Journal of Psychology, 217*(1), 49-60.

Mustonen, T., Berg, M., Kaistinen, J., Kawai, T., & Hakkinen, J. (2013). Visual task performance using a monocular see-through head-mounted display (HMD) while walking. *Journal of Experimental Psychology: Applied, 19*(4), 333-344. doi:10.1037/a0034635

Nikolic, M. I., Orr, J. M., & Sarter, N. B. (2004). Why pilots miss the green box: How display context undermines attention capture. *International Journal of Aviation Psychology, 14*(1), 39-52. doi:DOI 10.1207/s15327108ijap1401_3

Olzak, L. A., & Thomas, J. P. (1986). Seeing Spatial Patterns. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance* (Vol. 1, pp. 56). USA: John Wiley & Sons Inc.

Pascale, M., Sanderson, P., Liu, D., Mohamed, I., Stigter, N., & Loeb, R. (2015). *Peripheral detection for abrupt onset stimuli presented via head-worn display.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Patterson, R., Winterbottom, M. D., Pierce, B. J., & Fox, R. (2007). Binocular rivalry and head-worn displays. *Human Factors, 49*(6), 1083-1096. doi:10.1518/001872007x249947

Ruskin, K. J., & Hueske-Kraus, D. (2015). Alarm fatigue: impacts on patient safety. *Current Opinion in Anesthesiology, 28*(6), 685-690.

Sawyer, B. D., Finomore, V. S., Calvo, A. A., & Hancock, P. A. (2014). Google glass: A driver distraction cause or cure? *Human Factors, 56*(7), 1307-1321. doi:10.1177/0018720814555723

Sheridan, T. B. (1970). On how often the supervisor should sample. *Systems Science and Cybernetics, IEEE Transactions on, 6*(2), 140-145.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception, 28*(9), 1059-1074. doi:Doi 10.1068/P2952

Theis, S., Mertens, A., & Wille, M. (2015). *Effects of data glasses on human workload and performance during assembly and disassembly tasks.* Paper presented at the Proceedings of the 19th triennial congress of the IEA, Melbourne.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 50*(3), 449-455.

Wickens, C. D., Dixon, S. R., & Seppelt, B. (2005). *Auditory preemption versus multiple resources: Who wins in interruption management?* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Winterbottom, M. D., Patterson, R., Pierce, B. J., Covas, C. M., & Winner, J. (2007). Depth of focus and visual recognition of imagery presented on simultaneously viewed displays: implications for head-mounted displays. *Human Factors, 49*(5), 907-919. doi:10.1518/001872007x230253

Winterbottom, M. D., Patterson, R., Pierce, B. J., Gaska, J., & Hadley, S. (2015). *Visibility of monocular symbology in transparent head-mounted display applications.* Paper presented at the SPIE Defense+ Security.

Wolfe, J. M., O'Neill, P., & Bennett, S. C. (1998). Why are there eccentricity effects in visual search? Visual and attentional hypotheses. *Perception and Psychophysics, 60*(1), 140-156.

Woodham, A., Billinghurst, M., & Helton, W. S. (2016). Climbing With a Head-Mounted Display: Dual-Task Costs. *Human Factors, 58*(3), 452-461. doi:10.1177/0018720815623431

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance, 10*(5), 601-621.

Appendix A

Calibration Study

The stimuli on both displays needed to be modified so that the colors and contrasts on Google Glass matched that of the simulated HWD, and vice versa. Specifically, the background of the simulated HWD on the computer screen, using red, green, and blue (RGB) levels as well as opacity (transparency). Then the brightness of the bar stimuli on Google Glass (dark gray, light gray, and white) were modified to match those presented on the newly defined simulated HWD background by adjusting the intensity of the “whiteness”.

Participants (*n* = 11) wore Google Glass throughout the entire procedure and started at two different RGB values from which they adjusted the colors and opacity of each stimulus. One condition began with all color and contrast values set well above the expected range (Red: 255, Green: 240, Blue: 240, Opacity: 90, Stimuli: 150), while the other began with all values well below the expected range (Red: 255, Green: 150, Blue: 150, Opacity: 180, Stimuli: 30). These blocks were counterbalanced so that the starting values of the first block alternated across participants. During the second block, the values were hidden so that participants could not simply match the values they arrived at in the first block.

The resulting RGB and opacity values for the background of the simulated HWD were 255, 204, 182, and 72 respectively. These were the averages from both blocks of adjustments. Similarly, for the dark gray, light gray, and white bars on Google Glass, the values were 45, 87, and 113 respectively. For the color of the bars, the value (e.g. 45) was used for all three RGB values (e.g R: 45, G: 45, B: 45) and the opacity was locked at the maximum. Participants’ adjustments resulted in a simulated HWD background that more closely resembled a light rust color rather than black, which closely resembled what is visible on Google Glass, under the controlled lighting conditions in the testing environment. In other settings with different light, the Google Glass background, if a static color, would probably be very different from the results presented here. If used in a dynamic environment, the Google Glass background would be ever-changing.

Appendix B: Manuscript 2

Pascale, M., Sanderson, P., Liu, D., Mohamed, I., Brecknell, B., Loeb, R. (2017). Impact of head-worn displays on strategic alarm management. *Human Factors. Revise and resubmit, 7 August 2017.*

Impact of head-worn displays on strategic alarm management and situation awareness

Michael Pascale[1]
Penelope Sanderson[1,2,3]
Dave Liu[2]
Ismail Mohamed[1]
Birgit Brecknell[1]
Robert G. Loeb[1,4]

[1]School of Psychology, The University of Queensland, Australia
[2]School of ITEE, The University of Queensland, Australia
[3]School of Medicine, The University of Queensland, Queensland, Australia
[4]College of Medicine, University of Florida, Gainesville, Florida, USA

Running head: HWDs and patient monitoring
Manuscript type: Extended Multi-Phase Study
Abstract: 250 words
Word Count: 7226 words

Corresponding author: Michael T. Pascale – m.pascale@uq.edu.au

Abstract

Objective: To investigate whether head-worn displays help mobile participants make better alarm management decisions and achieve better situation awareness while performing an ongoing task.

Background: Patient alarms occur frequently in hospitals, but they often do not require clinical intervention. Over time, clinicians can become desensitized to alarms and may fail to respond to clinically relevant alarms. Head-worn displays can make patient information continuously accessible, support situation awareness, and help clinicians make informed decisions based on patients' status.

Method: Experiment 1 tested whether non-clinicians benefit from patient information that is continuously displayed on a head-worn display while the participant performs a secondary calculation task. Experiment 2 tested how participants with clinical experience monitored simulated patients for changes while in three different display conditions (three separate sessions, each separated by at least a week), and while they performed a complex simulated patient assessment.

Results: Experiment 1 showed that participants left patients in an untreated state for less time, and less often, when participants had access to continuous patient information on a head-worn display. Experiment 2 showed that participants using a head-worn visual and auditory display could more accurately describe the status of their monitored patients, but only in the first session.

Conclusion: Head-worn displays may help individuals maintain awareness of the statuses of multiple remote processes while simultaneously carrying out other tasks.

Application: The outcomes of the current study may apply to contexts where access to continuous streams of information from remote locations is useful, such as patient monitoring or clinical supervision.

Keywords: head-worn displays, divided attention, monitoring, alarms

Précis: Head-worn displays offer the ability to present continuously available, real-time patient information. In a computerized microworld environment, the HWD afforded participants the ability to treat patients faster, and more often, than participants without the display. Additionally, the HWD was associated with an increase in situation awareness and improvements to ongoing task performance.

Introduction

Physiological monitoring systems are used in hospitals to inform clinicians—doctors or nurses—about a patient's vital signs. However, monitoring displays are typically only available at a patient's bedside, at a nurses' station where a central display shows information from multiple patients, or more recently, on tablet computers. Centralized monitors and continuous patient monitoring can improve patient safety (Watkins, Whisman, & Booker, 2016) but, unfortunately, clinicians spend little time at the nurses' station, given the variety of tasks they need to perform. As a result, any alarms indicating that a patient's vital signs are deviating from a normal state need to reach the clinician, either through sounds that can be heard at a distance, marquee LED signs, pagers, or displays that can be accessed ubiquitously (Sanderson, 2006). Unfortunately, this generates a stimulus-saturated environment that can disrupt patient care, rather than improve it (Xiao, Seagull, Nieves-Khouw, Barczak, & Perkins, 2004). Head worn displays (HWDs), however, may help clinicians maintain awareness of their patients' well-being by displaying patient vital signs that are continuously available.

The purpose of the research reported in this paper is to test the usefulness of displaying continuous patient information on an HWD. In the remainder of the introduction, we provide background information on patient alarms and mobile monitoring devices, including HWDs. We then present the experimental microworld we used to examine the advantages and disadvantages of HWDs, compared with conventional alarms. Finally, we present the results of two studies.

The alarm problem and information accessibility

Patient alarms occur frequently in many hospital wards, but they seldom provide information about the events that triggered them. Clinicians need to decide, often blindly, whether an alarm is clinically relevant and whether a response is necessary. Unfortunately, between 64 and 99 percent of patient alarms can be categorized as clinically irrelevant, requiring no intervention (Paine et al., 2016; Ruskin & Hueske-Kraus, 2015). When uninformative alarms are combined with high rates of irrelevant alarms, clinicians can become desensitized, which results in alarms being shut off, ignored or missed (Paine et al., 2016). In most cases the alarms are not directly or immediately relevant to the patient's well-being, and therefore do not need immediate attention. Occasionally, however, patients with a clinically relevant issue are left unattended, sometimes resulting in their deterioration and, in a few notable cases, death (Xiao et al., 2004).

Efforts have been made to reduce the frequency of irrelevant alarms, but no technique has been universally adopted. Studies have tested the effectiveness of training clinicians to set more appropriate alarm thresholds (Graham & Cvach, 2010). Further studies have tested better algorithms for triggering alarms, such as using delays before alarm onset (Gorges, Markewitz, & Westenskow, 2009), and taking patient history into account (Imhoff, Kuhls, Gather, & Fried, 2009). However, the patient information being conveyed by the remaining auditory alarms is still limited and does not convey context, so there is still the concern that clinically important information may be missed.

A more effective solution may be to enhance clinicians' ease of access to information—for example, by providing a wearable display, such as an HWD, that directly accesses vital information from all the patients in a ward. The display would be comparable to the displays in a centralized nurses' station, or would present vital signs of all patients under that clinician's direct care. As long as the information on an HWD does not occlude vision, add annoyance, or cause distraction, clinicians may be able to respond to critical patient events in a more timely fashion, and even to preempt some problems. Additionally, clinicians could monitor their patients' levels of comfort, or any concerns that they have regarding patients. Similarly, access to continuous patient information would be useful for experienced team leaders who are providing support to several bedside nurses, where some of the latter may be relatively inexperienced or may have to handle high levels of workload.

Mobile monitoring and notification devices

Handheld devices have had some success in hospital environments, but they can introduce problems. For instance, pagers (Cvach, Frank, Doyle, & Stevens, 2014) and tablets (Baig, GholamHosseini, & Lindén, 2015) have improved response times to some patient events, but they can introduce additional contagion hazards because clinicians need to physically interact with them to get information.

HWDs have been trialed in several professional domains but their adoption is not yet widespread. Within healthcare, HWDs have been explored as a way to supplement anesthesiologists' computer monitors (Liu, Jenkins, & Sanderson, 2009b), allowing doctors to spend more time focusing directly on patients than on monitoring equipment, and, in at least one reported instance, to detect an unexpected patient event that may have otherwise gone unnoticed (Vorraber et al., 2014). In other domains, HWDs and head-up displays have helped pilots to detect events (Ververs & Wickens, 1998) and to process multiple streams of visual information (Wickens & Long, 1995).

HWDs, however, are sometimes associated with visual, perceptual, and performance costs (Wickens, Goh, Helleberg, Horrey, & Talleur, 2003). For example, stimuli presented on an HWD can go undetected (Liu, Jenkins, Sanderson, et al., 2009; Pascale et al., 2015), can have an impact performance on ongoing tasks (He, Ellis, Choi, & Wang, 2015; Sawyer, Finomore, Calvo, & Hancock, 2014; Woodham, Billinghurst, & Helton, 2016), and can reduce the user's ability to walk normally (Mustonen, Berg, Kaistinen, Kawai, & Hakkinen, 2013).

The primary limitation with regards to using an HWD is that it draws on visual resources that might otherwise be allocated towards other parts of the visual field. Attention is a limited resource that is strongly absorbed by perceptually engaging tasks (Lavie, 2005; Lavie, Beck, & Konstantinou, 2014; Raveh & Lavie, 2015). When a person allocates more attentional resources to a task, the locus of their attention appears to narrow, reducing their ability to notice non-task stimuli. Information that is displayed on an HWD can go unnoticed if the real environment demands constant visual attention. Conversely, information in the environment might be missed if the display on the HWD draws too much visual attention.

Despite the above concerns, the advantages of having information continuously available to the clinician may compensate for the potential disadvantages associated with HWDs, and more importantly, the disadvantages of having no information at all. In other words, an HWD of vital signs does not require continuous monitoring. It could be designed so that it does not disrupt vision or attention, and could let a clinician access patient information with only a few saccades when the information might otherwise be difficult, impractical, or even impossible to access without moving to a patient's bedside or to a stationary centralized monitoring system.

Description of the patient monitoring microworld

Our goal was to test the effectiveness of HWDs for improving performance relevant for patient well-being in a simple microworld simulation of a hospital ward. We designed the computerized microworld to capture a specific challenge faced by clinicians monitoring multiple patients: the separation in space between clinician and patient, and therefore the temporal delay for the clinician to move bedside to see details of a patient's vital signs unless the clinician uses a portable technology. The purpose of Experiment 1 was to test untrained participants' ability to use continuous streams of vital sign information to prioritize their attention towards patients experiencing clinically relevant alarms, as opposed to irrelevant (sensor) alarms, in the simulated ward. In addition to responding to alarms, participants performed an ongoing task located separately from the microworld, requiring them to stand and walk between two different task locations. In Experiment 2, we tested nursing students, introduced an additional auditory "notification" condition to replace bedside alarms, and increased the complexity of the tasks.

Furthermore, Experiment 2 was designed to measure participants' situation awareness (Endsley, 1995), operationalized using the Situation Awareness Global Assessment Technique (SAGAT), rather than their ability to use continuous displays to prioritize alarms of differing relevance.

The layout of the experiment can be seen in Figure 1. Participants completed the ongoing tasks and the monitoring tasks at computer stations on opposite sides of a moderately-sized room. The microworld monitoring station was enclosed on three sides by opaque curtains, so that the display for one task could not be seen from the other task. To access the patients' bedsides, participants had to stand up and walk from the ongoing task station to the monitoring station, and from there use the computer's mouse to navigate between virtual rooms in the patient monitoring microworld. For Experiment 2 there was an additional computer at the ongoing task station that was used to display the SAGAT questions. Participants leaned over to the second computer to answer the questions during the SAGAT freezes.

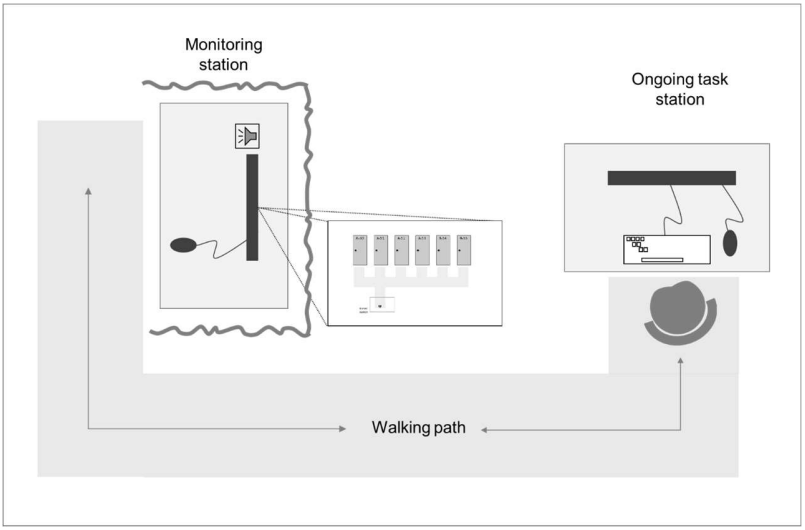

Figure 1. Diagram of the layout for both experiiments. For Experiment 2 an additional computer was located at the ongoing task station that showed the SAGAT questions during freezes.

The microworld consisted of a set of six patients behind six room doors (see Figure 2). Each patient had vital signs that were displayed when the room was entered, as might be seen when moving to the patient's bedside. For the current experiments, we selected three of the most commonly measured vital signs: heart rate (HR), blood pressure (BP), and oxygen saturation (SpO_2). The time it took to move between simulated rooms approximated the time it might take to walk along a physical corridor: from 5.6 seconds to move between adjacent patients, to 12 seconds to move from the leftmost patient to the rightmost patient.

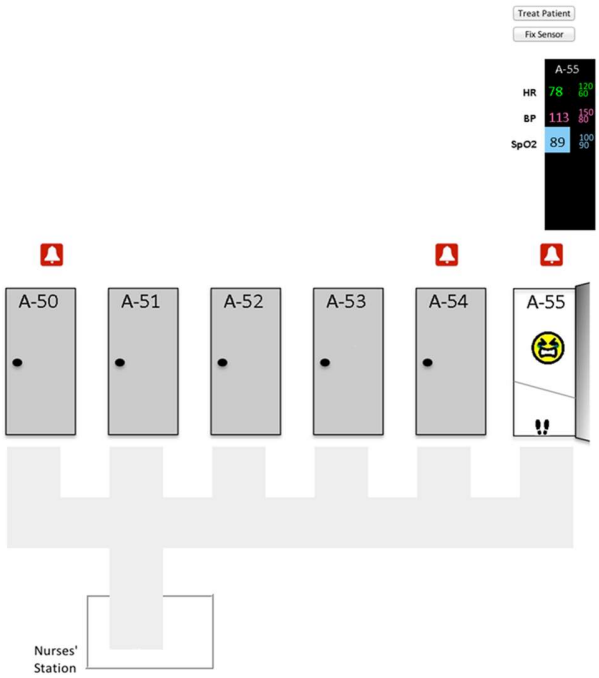

Figure 2. Example of the participant's view of the microworld in the alarms only condition. The screenshot shows three active alarms (see red icons above doors). The participant is inside the sixth patient's room (door is open and footprints icon is in the A-55 doorway) and therefore sees vital signs for just that patient (placed above the relevant door). The patient in A-55 is experiencing low oxygen saturation (89%), which triggered the alarm.

We included two types of alarms to recreate the situation described in the introduction when, more often than not, alarms are not actionable by the clinical staff. The first type of alarm, a 'clinical alarm', was triggered if the value of a vital sign trended into a high priority range, indicating the start of a clinically relevant problem. For example, a heart rate might increase steadily from 80 to 120 bpm, and the alarm would be triggered at 120bpm. The second type of alarm, a 'sensor' alarm, reproduced what might happen if a physiological sensor become disconnected or failed. Specifically, a value in the normal range changed instantly to either zero or 300. For example, a displayed oxygen saturation level might change suddenly from 99% to 0%, while the patient's true oxygen saturation remained at 99%. A sensor alarm therefore had no immediate clinical relevance. Both alarm types triggered the same visual and audible alarms. The alarms sounded from a speaker at the microworld station, and sounded every 15 seconds while the vital sign was outside of the normal range. The alarm logic and alarm sounds are described more fully in each experiment and in Appendix A.

Alarms were also displayed visually as red icons above a patient's door, as shown in Figure 2. The alarms were also displayed on the HWD and bedside displays (if the participant was in that patient's room), by inverting the colors of the background and font for that vital sign (see Figure 3). For example, SpO_2 is normally written in blue on a black background, but in an alarm state it was written in black on a blue background, as can be seen in Figure 2 and Figure 3.

Figure 3. Example of the participant's view of the continuous streams of information on the HWD in Experiment 1 (left) and Experiment 2 (right).

Experiment 1

The purpose of Experiment 1 was to investigate whether access to a continuous information display, presented via an HWD, improved participants' ability to prioritize the treatment of sick patients in the microworld while they performed an ongoing task. As noted in the Introduction, HWDs can interfere with cognition and perception by affecting participants' ability to notice stimuli. Moreover, an ongoing task that absorbs attentional resources will affect participants' ability to monitor a continuous information display (Lavie, 2005), or vice versa. Even so, participants using the HWD should make better decisions than participants not using the HWD about when to suspend performance on the ongoing task and walk to another physical location in the test room to handle clinically relevant alarms.

The participants' goal was to complete an ongoing task (dosage calculations) at the ongoing task station, and manage patient alarms at the monitoring station (see Appendix A). Participants were instructed that they should prioritize clinical alarms by treating patients. If no clinical alarms were present, they should complete the dosage calculations. We hypothesized the following.

(1) Participants in the HWD plus alarms (HWD-alarms) condition would treat patients faster than participants in the alarms only condition.

(2) In the low frequency condition there would be no difference in treatment latency between participants in the HWD-alarms and alarms only condition, but in the high-frequency

condition participants in the HWD-alarms condition would treat alarms faster than participants in the alarms only condition.

(3) Participants in the HWD-alarms condition would spend more time performing the ongoing task (drug dosage calculation), than participants in the alarms only condition, regardless of alarm frequency.

(4) Participants in the HWD-alarms condition would report lower workload than participants in the alarms only condition.

Method

Participants. The participants (n=76) were students at The University of Queensland who participated for either course credit or $20 gift cards. Exclusion criteria were the use of corrective eyeglasses or low levels of English comprehension. If the experimenter found that a participant had difficulty understanding the instructions, they were excluded from participation prior to any data being collected from them. The research complied with the Australian National Statement on Ethical Conduct in Human Research and was approved by the Human Research Ethics Committee of the School of Psychology at The University of Queensland (approval 14-PSYCH-PHD-59-JS). Informed consent was obtained from each participant.

Design. Experiment 1 was a 2 x 2 between-subjects design with factors of display condition (HWD-alarms versus alarms only) and alarm frequency (low versus high). In the HWD-alarms condition participants saw the continuous information display on a monocular see-through HWD (see left of Figure 1). In the alarms only condition, participants did not see the continuous information display and could only see a specific patient's vital signs when they were inside that patient's room (see Figure 2 for an example). In the low alarm frequency condition there were 22 alarms per block (seven clinical and 15 sensor) and in the high frequency condition 45 alarms per block (13 clinical and 32 sensor).

Outcome variables. We measured several different performance variables throughout each block. These were: (a) treatment latency, or the latency to treat clinical alarms (milliseconds), (b) untreated alarms, or the number of clinical alarms left untreated (count), (c) dosage calculation time, or the cumulative amount of time answering dosage questions at the dosage task computer (minutes), (d) dosage accuracy, or the number of ongoing task questions attempted (count) and answered correctly (count), and (e) subjective workload, or responses to each item on the NASA-TLX (0 to 21, low to high).

Apparatus. Auditory alarms were three successive tones that alerted the participant when a patient entered a high priority state. The tones used were the IEC 60101-1-8 General alarm, medium priority (IEC60601-1-8, 2005– 08). One instance of the alarm sound occurred for

each new alarm, and it repeated every 15 seconds until that alarm was either managed by the participant or shut off by the software, explained as a 'simulated nursing aide.' The clinical alarm states always lasted for 30 seconds before being shut off, but the sensor alarms lasted anywhere from 11 to 72 seconds (*M* = 34.9). Appendix A contains the ranges that were used for each vital sign, as well as a more detailed description of how the alarms functioned.

The HWD used in Experiment 1 was Google Glass (Google Inc., Mountain View, CA) which uses a monocular optical see-through display. The HWD was connected wirelessly to the computer running the microworld and the values on the display updated every two seconds. In the alarms only conditions, participants wore the HWD as a control measure, but no information was displayed on its screen. In the HWD-alarms conditions, participants wore the HWD which displayed a continuous stream of vital signs from all six patients.

Tasks. The ongoing task consisted of a series of "dosage calculation" problems that required multiplication and division of whole numbers (Figure 4) and a true/false response. Participants responded at their own pace using the keyboard. Participants were asked "to perform as many dosage calculations as you can, while keeping your simulated patients healthy."

For the monitoring task, participants responded to patient alarms by clicking on a patient's room which triggered the footsteps to walk to the room. The participant then decided either to treat the patient or fix the sensor, using either the "Treat Patient" or "Fix Sensor" button above the vital sign information (vital signs could not be fixed until they transitioned into an alarm range). Participants were instructed to prioritize sick patients (indicated by clinical alarms) over clinically irrelevant sensor failures (indicated by sensor alarms). Participants were also told that they could "safely ignore the sensor alarms to complete dosage calculations."

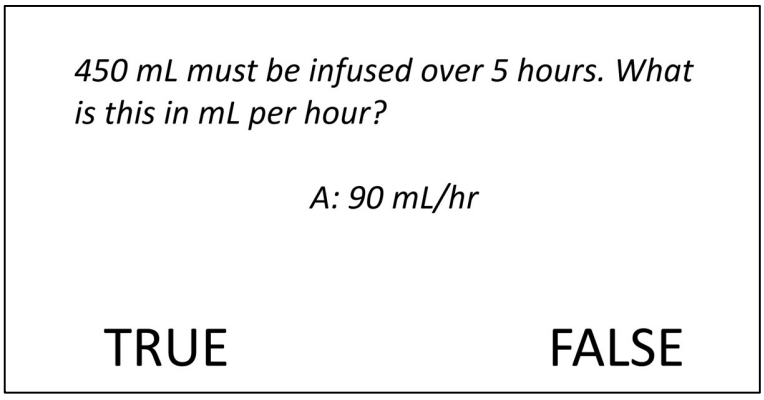

Figure 4. An example dosage calculation as seen by participants. Responses were entered using keys on the laptop keyboard to indicate whether the answer was true or false. Immediate feedback was given following a response and the next trial was shown after two seconds.

Procedure. Participants saw recorded PowerPoint instructions and then performed practice and experimental blocks that were relevant for their experimental condition. The instructions explained how the microworld was set up, how to move between rooms, what the three vital signs signified, and how the alarms functioned. The instructions included specific descriptions of the normal, trending, and critical ranges for each vital sign. Participants were told that any vital signs that were in the trending range, would eventually trigger a critical alarm. They were also informed that treating critical alarms was their top priority. The ranges were posted at each computer station for them to reference through the experiment. Participants in the HWD-alarms condition saw additional instructions regarding the vital sign display on Google Glass. All participants then completed a one-minute walkthrough and a six-minute practice block. After completing the practice block, they completed three 10-minute scenarios of patient monitoring in their condition. They completed the NASA-TLX after the first and third scenarios.

Analyses. Analyses were conducted on 76 participants, with 19 participants in each of the four conditions. We conducted a series of regressions to explore the effects of display condition and alarm frequency. Initial models always contained the full interaction. Any further models are described below.

- The first analysis was a multiple linear regression testing the interaction of the two independent variables on treatment latency, tested against a Bonferroni-adjusted p-value of .0167 to preserve an alpha of .05. For this analysis, all the treated clinical alarms were averaged for each participant (excluding missed alarms) and transformed to normality using a logarithmic function.
- The second analysis was a multiple linear regression to test the effects of display condition and alarm frequency on the dosage calculation time. Each participant's time was transformed from seconds to minutes.
- A third multiple linear regression tested the effects of the display condition and alarm condition on subjective workload. The ratings for the six subscales of the NASA-TLX were averaged to arrive at an unweighted workload score (0–21) for each participant.
- A fourth regression was conducted to test the effects of display condition and alarm frequency on the number of untreated alarms. The data for untreated alarms fit a Poisson distribution, so we conducted a Poisson regression.
- A final multiple linear regression tested the effects of display condition and alarm frequency on dosage calculation accuracy.

Results

Descriptive statistics can be found in Table 1. A summary of all regression analyses can be found in Table 2. The regression analyses were used to generate models, where possible, that predict each outcome variable (see Figure 5).

The results of the first regression showed that the average latency to treat clinical alarms was significantly shorter in the HWD-alarms condition than in the alarms only condition. The average latency to treat clinical alarms was significantly shorter in the low alarm frequency condition than in the high alarm frequency condition. Analyses did not reveal a significant interaction between display condition and alarm condition (p = .419).

Table 1

Observed means and standard deviations for the dependent variables in Experiment 1

	Treatment latency		Dosage time		Dosage accuracy		Untreated alarms		Subjective workload	
Display condition	Low	High	Low	High	Low	High	Low	High	Low	High
Alarms only	15.71 (4.37)	19.46 (3.54)	13.27 (2.42)	6.56 (2.08)	86.03 (6.65)	80.13 (9.46)	1.89 (2.54)	7.16 (7.13)	9.77 (2.48)	10.86 (3.48)
HWD-alarms	11.95 (1.45)	14.12 (2.3)	14.58 (2.99)	9.45 (3.66)	83.64 (6.89)	81.97 (7.39)	0.05 (0.23)	1.32 (3.97)	8.8 (2.72)	10.32 (3.16)

Table 2

Summary of the five regression analyses in Experiment 1. Variables show difference from the baseline condition of alarms only, low frequency of alarms.

	Treatment latency			Dosage time			Subjective workload			Untreated alarms			Dosage Accuracy		
Variable	*β*	*SE*	*t*	*β*	*SE*	*t*	*β*	*SE*	*t*	*β*	*SE*	*t*	*β*	*SE*	*t*
Intercept	2.72	0.04	62.16	13.27	0.65	20.28	9.77	0.68	14.27	0.64	0.17	3.83	86.03	1.76	48.82
HWD-alarms	-0.24	0.06	-3.96**	1.32	0.93	1.42	-0.97	0.97	-1.00	-3.58	1.01	-3.53**	-2.39	2.49	-0.96
High Alarms	0.23	0.06	3.79**	-6.71	0.93	-7.25**	1.08	0.97	1.12	1.33	0.19	7.09**	-5.90	2.49	-2.37
HWD-alarms x High Alarms	-0.07	0.09	-0.81	1.58	1.31	1.20	0.43	1.37	0.31	1.89	1.04	1.82	4.23	3.52	1.20

Note. All *df* = (3,72). ** $p < .001$.

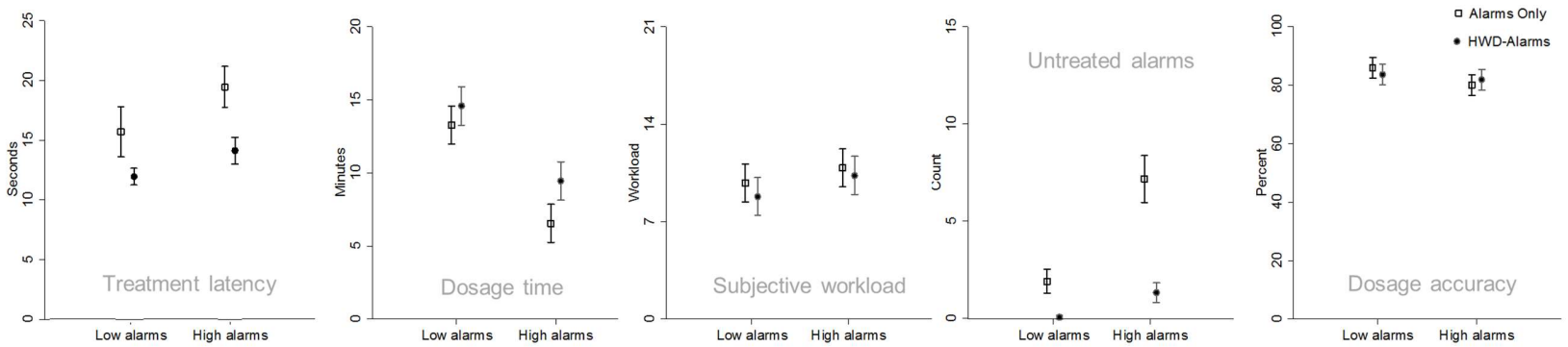

Figure 5. Predicted values and 95% confidence intervals for each outcome variable. Treatment latency could not generate a figure with the same level of clarity after being transformed, so the observed values and 95% confidence intervals were used to generate the figure (far left).

Results of the second regression on dosage calculation times showed that when the alarm frequency was low, participants in the HWD-alarms condition did not spend more time on the calculations than participants in the alarms only condition. The regression did show that participants spent less time on the dosage calculations in the high alarm frequency condition than in the low alarm frequency condition. There was no interaction (p = .233). However, a planned comparison revealed a significant difference between the display conditions when alarm frequency was high (SE = .93, t = 3.12, p = 0.003).

The final three regressions yielded mixed results. The regression on subjective workload revealed no differences between any conditions. The regression on untreated alarms showed that participants in the HWD-alarms condition relied less on the simulated nursing aide to treat clinical patients than did participants in the alarms only condition. Lastly, there were no differences between participants' dosage calculation accuracy in the HWD-alarms condition versus the alarms only condition, or between low and high frequency alarm conditions.

Discussion

As hypothesized (H1), participants in the HWD-alarms condition, who had access to continuous patient information, treated clinical alarms faster on average than did participants in the alarms only condition, who did not have continuous access to patient information. The continuous display improved prioritization and decision making when participants were faced with multiple alarms, but regardless of how frequently alarms occurred (against H2). H3 was partially supported, participants in the HWD-alarms condition performed the ongoing task for longer, but only in the high alarms condition. H4 was not supported: participants' subjective workload in the alarms only condition was not significantly different from participants in the HWD-alarms condition.

Further analyses revealed that participants in the alarms only condition left more clinical alarms untreated than participants in the HWD-alarms condition, regardless of alarm frequency, probably because participants in the HWD-alarms condition could identify clinical alarms without having to move from room to room to "find" the truly sick patients. Taken together, these results provide evidence that using a continuous display of patient vital signs for multiple patient monitoring can improve decision making and prioritization behaviors, supporting the findings from Watkins et al. (2016).

Experiment 2

The data from the Experiment 1 showed that participants can use continuous information displayed on an HWD to make better task prioritization decisions in our simulated microworld. The next goal was to explore whether the continuous information increases situation awareness.

Situation awareness has been highlighted as fundamental to nurses' decision-making (Stubbings, Chaboyer, & McMurray, 2012) and their management of deteriorating patients (Bogossian et al., 2014), both of which can be enhanced using advanced, stationary displays (Koch et al., 2013; Zhang et al., 2002). We investigated whether mobile displays could also improve situation awareness for the status of multiple patients.

In Experiment 2, we tested how participants with clinical experience monitored six simulated patients for physiological changes using three display conditions—alarms only, HWD-alarms, and HWD-notifications). HWD-notifications introduced an alternative auditory display that intermittently played individual sounds for each patient, in order, from the HWD's earpiece. We hypothesized that this "notification" display might be more informative than standard bedside alarms. Additionally, we replaced the dosage calculation task with a more complex simulated patient assessment task. Participants' goal was to complete as much of the patient assessment as they could, but to also maintain awareness of a further set of six patients. We hypothesized the following:

H1 Participants would answer situation awareness questions more accurately in each HWD condition than in the alarms only conditions.

H2 Participants would answer situation awareness questions more accurately in the HWD-notifications condition than in the HWD-alarms condition.

H3 Participants would obtain higher scores on the ongoing patient assessment task in each of the HWD conditions than in the alarms only condition.

H4 Participants would report lower workload in the HWD conditions than in the alarms only condition.

H5 Participants would report that the alarm sounds are "more annoying" than the notifications.

Method

Participants. To be eligible for participation, students had to be enrolled in The School of Nursing and Midwifery, and have completed their first year. Our aim was to achieve a total sample size of 20 participants; however, due to competing curriculum demands this number was not achieved. Thirteen second and third year nursing students from The University of Queensland volunteered and received $20 per hour in gift cards. The research complied with the Australian National Statement on Ethical Conduct in Human Research and was approved by the Human Research Ethics Committee of the School of Psychology at The University of Queensland (approval 14-PSYCH-PHD-59-JS). Informed consent was obtained from each participant before each session.

Design. Experiment 2 was a within-subjects study. Each participant's situation awareness of patients' status was measured across three different display conditions, while the participant performed a simulated patient assessment. The first two display conditions were identical to those in Experiment 1 (alarms only versus HWD-alarms). In the third condition, the HWD-notifications condition, the alarms were replaced with a sequential auditory display that played individual sound motifs for each patient, in the order they were displayed left-to-right on the HWD. Participants performed with each of the display conditions (counterbalanced) across three sessions, each session separated by at least six days.

Outcome variables. We measured (a) SAGAT accuracy, or the number of correct responses to the SAGAT questions (count), (b) vSim score, or performance on the ongoing task (percent completed), (c) subjective workload, or responses to each item on the NASA-TLX (0 to 21, low to high), (d) sound annoyance, in answer to the question "How annoyed were you by the sounds used to convey patient information?" (0 to 21, annoying to not annoying), (e) device use, in answer to the question "How easy was it to use the head-worn display to monitor your patients?" (0 to 21, not easy to very easy), and (f) display use, in answer to the question "How easy to use was the layout of the vital sign display?" (0-21, not easy to very easy). Questions about HWD device use and display use were not included in the alarms only condition.

Apparatus. Participants were instructed to maintain awareness of their patients and keep track of changes to their status. For this experiment, there were two possibilities for alarms. The first, a warning alarm, suggested that a patient had at least one vital sign outside of the normal range (IEC General medium), and the second, a critical alarm, suggested that a patient had at least one vital sign that was critically high or critically low (IEC General high). Alarms were triggered whenever a vital sign crossed a threshold outside of the normal range, e.g. normal to warning, warning to critical, or critical to warning, but not warning to normal. The nonactionable technical alarms always generated a critical alarm sound. The alarm ranges were posted at each computer station for the participant to reference throughout the experiment. The ranges and a fuller description of the alarms can be found in Appendix A and Table A2.

In the HWD-notifications condition, the auditory display sounded from the earpiece of the HWD, and was triggered any time any threshold was crossed, including critical to warning, and warning to normal. The notifications consisted of six consecutive sounds that conveyed the status of each of the six patients, in the same order as the visual display. The sound played for each patient was based on their status, which was conveyed using one of three 500 ms tones. A low-pitched beep indicated that all of a patient's vital signs were normal. A medium-pitched beep with slow vibrato (tremolo) indicated that at least one vital sign was in the warning range. A high-pitched beep with faster tremolo indicated that at least one vital sign was in the critical range. The beep

for each patient was separated from its neighbor by 400 ms. The total cycling display lasted five seconds. If any patient's status changed while the display was already playing, the audio was triggered again immediately afterwards, with the new status. A further description of the notification display can be found in Appendix A.

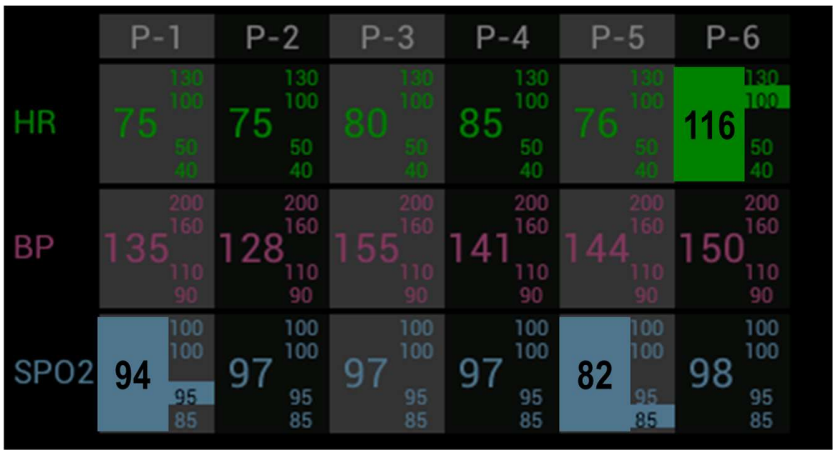

Figure 6. The layout of the vital sign display on the HWD.

The HWD in Experiment 2 was the Vuzix M100 (Vuzix, Rochester, NY, USA), with a monocular opaque LCD display. For this version of the display (see Figure 6), the patients were again laid out in columns, but we included the different ranges from the Q-ADDS (small values on the right side of each column). When a value was outside the normal range, the exact range was highlighted on the right, and the colors were inverted. For example, in Figure 8 patient 6 (P-6) has an HR greater than 100, which is in the warning range. Alternatively, patient 5 (P-5) has an SpO2 less than 85, in the critical range. The HWD display updated every 2 seconds.

Tasks. Participants monitored their patients for status changes and responded to SAGAT questions that probed perception of changes (SA1) and state comprehension (SA2). The questions occurred at fixed points (program freezes) that differed across blocks, so to the participant, the questions occurred at unpredictable intervals. There were two questions at each freeze, and a total of 10 questions per block. The questions asked about the status of a patient, or patient(s), or about recent changes to patient's status, such as "Which patient has the highest blood pressure," or "How many patients have low heart rate?" Participants responded by selecting the correct answer from a bank of two to four possible responses. During freezes, alarms and notifications were silenced, and the HMD went blank.

The simulated patient assessment task (vSim for Nursing, Laerdal) was completed on a separate computer. This task required participants to complete a patient assessment using mouse clicks (Figure 7). Each participant saw the same 6 unique patients, one per block, in the same order across sessions (except for one participant who saw the patients for session 2 in

session 1, due to an experimenter error). During the SAGAT freezes, the patient monitoring display was removed. Participants were instructed to immediately stop the assessment task, no matter what was happening to the patient, so that they could answer the questions. A score (0-100%) was generated automatically in the commercial software for the participant and depended on the total amount of the scenario that was completed, as well as the number, and severity of errors that were made.

Figure 7. Participant's view of the vSim for Nursing task. Actions could be completed using the options at the right. Each tab (colored rectangles to the right of vertical center) contained a complex set of options, often with further sub-options.

Procedure. In the first session, participants saw recorded PowerPoint instructions that described both tasks and all three display conditions. After the instructions, the participant received further instructions on the vSim for Nursing task, followed by two minutes of practice. In the second and third sessions, participants saw recorded PowerPoint instructions again, but only the slides that pertained to the display condition they would be seeing in that session. For all blocks, after the instructions, participants completed a five-minute practice block, then two 15-minute experimental blocks (counterbalanced across sessions), followed by one instance of the NASA-TLX questionnaire.

Analyses. We conducted a series of exploratory mixed effects linear regression analyses. For each regression, we included the interaction between display condition and session. Any further model fitting is described below.

- The first regression tested the effects of display condition and session on SAGAT accuracy. SAGAT accuracy was transformed to from a count to percentage correct by dividing the count of correct responses in each session into the total number of questions (20) per session.

- The second regression tested the effects of display condition and session on vSim scores.
- The third regression tested the effects of display condition and session on NASA-TLX responses. As in Experiment 1, the ratings for the six subscales of the NASA-TLX were averaged to arrive at an unweighted workload.
- The final three regressions tested the effects of display condition and session on sound annoyance, device use, and display use.

In three instances, the participant failed to stabilize the patient, resulting in a lower than normal patient assessment score, and early termination of the vSim task. In these instances, an additional patient was immediately loaded, but performance with that patient was excluded from the analyses.

Results

Descriptive statistics are reported in Table 3. Analyses were conducted on 13 participants who each saw all three display conditions. Due to the uneven number of participants, one display condition is over represented in each session (five observations instead of four).

The first regression, which analyzed SAGAT accuracy, showed that SAGAT accuracy in the HWD-alarms condition in session 1 was not significantly different from accuracy in the alarms only condition (see Table 4). However, SAGAT accuracy in the HWD-notifications was significantly greater than the alarms only condition in session 1. A planned comparison using session 1 data revealed a significant difference between the HWD-notifications condition and the HWD-alarms condition (β = 9.67, SE = 2.87, z = 3.37, p = .001). In sessions two and three, however, there were no differences between the display conditions. This regression model was used to produce the predictive accuracies for each condition (Figure 8).

The second regression model, which analyzed the vSim task performance data (Table 5, Model 1) revealed no significant differences between the three display conditions, in any session (Figure 9, left). A reduced model was fit to explore the differences between display conditions with the values for session at their means. This model (Table 5, Model 2) revealed a significant difference between the alarms only condition and each of the HWD conditions (Figure 9, right). There was no difference between the two HWD conditions (p = .339).

Table 3

Means and standard deviations for all dependent measures in Experiment 2

		SAGAT Accuracy	vSim Score	Subjective Workload	Sound Annoyance	Device Use	Display Use
Session 1							
	Alarms only	58.00 (2.74)	53.70 (8.21)	15.6 (2.61)	9.6 (5.13)	-	-
	HWD-Alarms	50.00 (14.72)	62.25 (6.98)	9.83 (3.82)	4.00 (6.00)	20.75 (0.5)	17.50 (4.36)
	HWD-Notifications	76.25 (14.36)	57.46 (17.29)	8.46 (3.14)	7.75 (7.41)	18.25 (2.5)	14.25 (2.22)
Session 2							
	Alarms only	67.50 (11.90)	44.00 (6.60)	14.58 (1.51)	9.75 (10.11)	-	-
	HWD-Alarms	78.00 (11.51)	63.8 (13.17)	9.37 (3.23)	6.40 (3.51)	16.8 (2.28)	17.8 (1.92)
	HWD-Notifications	71.25 (10.31)	58.38 (2.14)	8.25 (3.5)	1.50 (0.58)	19.75 (1.50)	18.5 (2.52)
Session 3							
	Alarms only	65.00 (10.80)	59.38 (6.76)	11.75 (3.38)	10.00 (6.16)	-	-
	HWD-Alarms	70.00 (18.71)	58.88 (6.96)	7.13 (1.85)	6.25 (4.72)	15.75 (5.74)	14.5 (3.42)
	HWD-Notifications	79.00 (8.22)	61.5 (10.73)	8.27 (2.30)	4.80 (4.92)	17.4 (4.16)	18.2 (3.90)
Total							
	Alarms only	63.50 (8.48)	52.36 (7.19)	13.98 (2.50)	9.78 (7.13)	-	-
	HWD-Alarms	66.00 (14.98)	61.64 (9.04)	8.78 (2.97)	5.55 (4.74)	17.77 (2.84)	16.6 (3.23)
	HWD-Notifications	75.50 (10.96)	59.11 (10.05)	8.33 (2.98)	4.68 (4.30)	18.47 (2.72)	16.98 (2.88)

Table 4

Summary of mixed effects linear regression for variables predicting SAGAT accuracy, against the baseline of alarms only in session 1

Variable	*β*	*SE*	*z*	*p*	95% CI
Intercept	56.07	4.25	13.19	< .01	47.74, 64.40
HWD-Alarms	-4.32	5.83	-0.74	0.46	-15.74, 7.11
HWD-Notifications	20.82	5.83	3.57	< .01	9.40, 32.25
Session-2	9.50	5.83	1.63	0.10	-1.93, 20.92
Session-3	13.26	5.83	2.27	0.02	1.83, 24.69
HWD-Alarms x Session-2	16.44	8.91	1.84	0.07	-1.03, 33.91
HWD-Alarms x Session-3	3.61	9.01	0.40	0.69	-14.05, 21.27
HWD-Notifications x Session-2	-12.83	9.01	-1.42	0.15	-30.49, 4.83
HWD-Notifications x Session-3	-13.53	8.91	-1.52	0.13	-31.00, 3.94
	Estimate	SE			
Random effects: Participant	60.56	32.57			21.10, 173.77

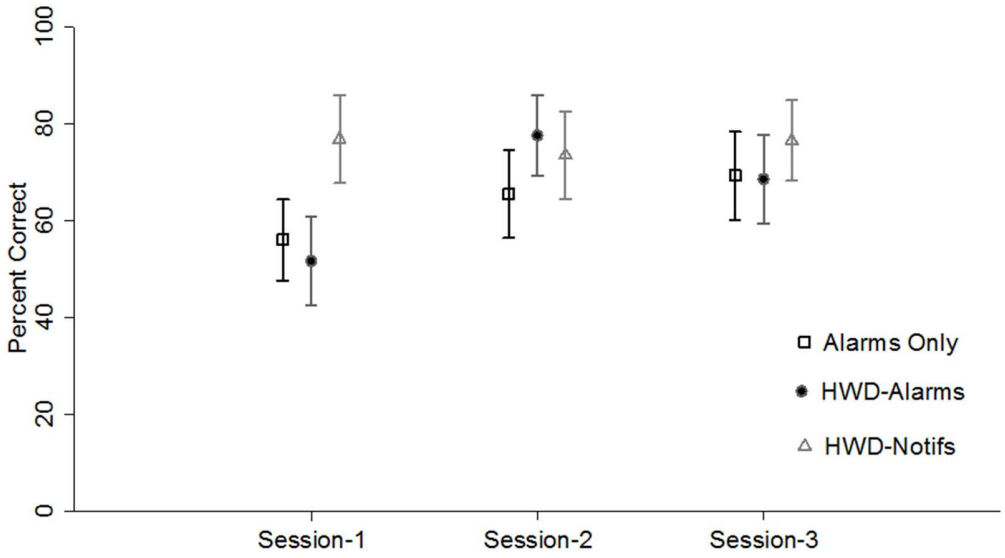

Figure 8. Predicted SAGAT accuracy and 95% confidence intervals for each display condition, across each session.

Table 5

Summary of mixed effects linear regression for variables predicting vSim score

	Model 1					Model 2				
Variable	*β*	*SE*	*z*	*p*	95% CI	*β*	*SE*	*z*	*p*	95% CI
Intercept	53.83	4.26	12.62	< .01	45.47 - 62.19	52.46	2.62	20.04	< .01	47.33 - 57.59
HWD-Alarms	8.32	6.29	1.32	0.19	-4.01 - 20.66	9.35	2.63	3.56	< .01	4.19 - 14.5
HWD-Notifications	3.22	6.29	0.51	0.61	-9.12 - 15.55	6.83	2.63	2.60	0.01	1.68 - 11.99
Session-2	-6.55	6.29	-1.04	0.30	-18.89 - 5.78					
Session-3	3.47	6.29	0.55	0.58	-8.86 - 15.81					
HWD-Alarms x Session-2	6.57	9.33	0.70	0.48	-11.71 - 24.86					
HWD-Alarms x Session-3	-4.63	9.51	-0.49	0.63	-23.27 - 14.01					
HWD-Notifications x Session-2	8.00	9.51	0.84	0.40	-10.64 - 26.64					
HWD-Notifications x Session-3	0.93	9.33	0.10	0.92	-17.36 - 19.22					
	Estimate	SE				Estimate	SE			
Random effects: Participant	23.28	21.56			3.79 - 143.02	44.18	23.57			15.53 - 125.71

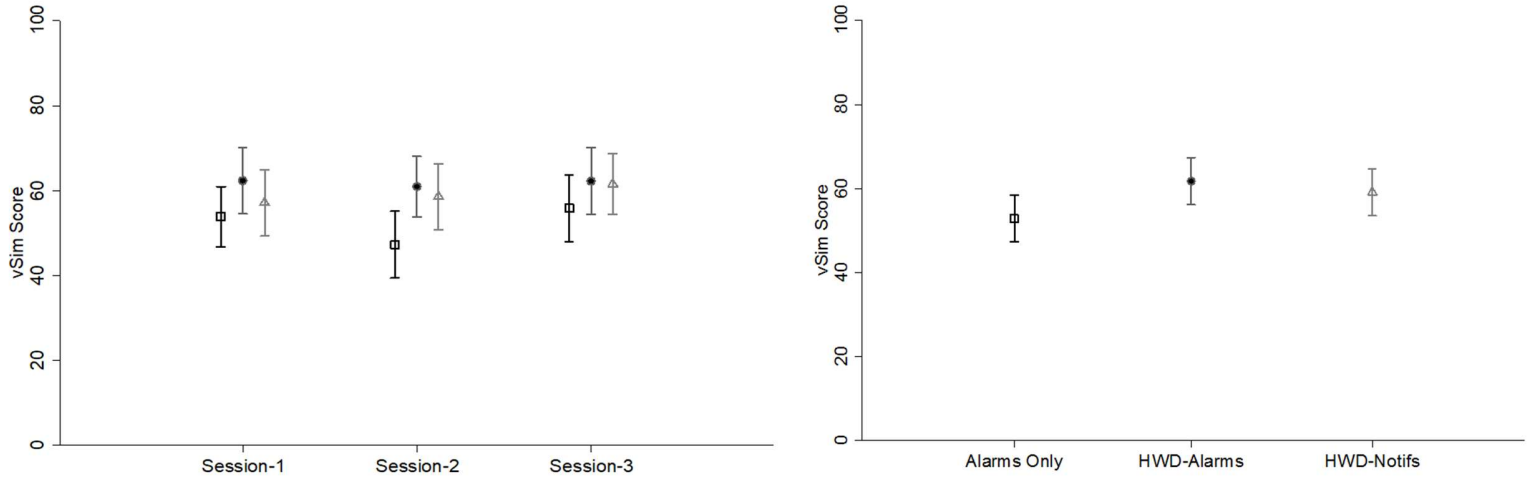

Figure 9. Predicted vSim scores and 95% confidence intervals for each of the regression models. Model 1 is on the left and Model 2 is on the right.

The third regression, which analyzed subjective workload, revealed a significant difference (in session 1) between the alarms only condition and the HWD-alarms condition (β = -6.26, SE = 1.32, z = -4.76, p < .001), and between the alarms only condition and the HWD-notifications condition (β = -6.45, SE = 1.32, z = -4.90, p < .001). There were no significant interactions between display condition and session. Further pairwise comparisons revealed that the difference between the alarms only condition and the two HWD conditions were significant in all three sessions (all p < .02). Figure 10 shows the predicted workload for each condition in each session.

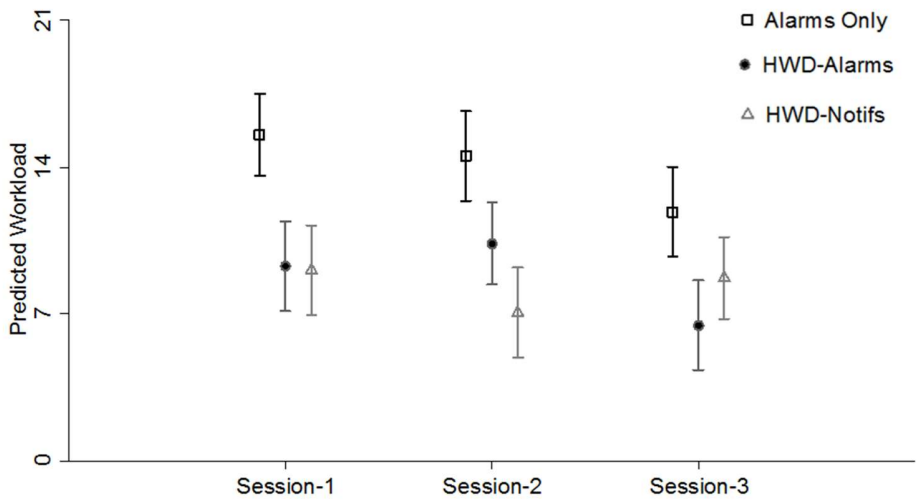

Figure 10. Predicted workload and 95% confidence intervals for each of the display conditions across sessions.

The initial regression analysis on sound annoyance in each condition resulted in no significant differences. After removing the interaction and focusing on the difference between display conditions, the regression indicated that participants found the sounds in the alarms only condition to be “more annoying” than the sounds in the HWD-alarms condition (β = -4.15, SE = 1.47, z = -2.82, p = .01) even though the same alarm sounds were used in each case. Similarly, participants reported that the alarms in the alarms only condition were significantly more annoying than the earcons used in the HWD-notifications condition (β = -5.08, SE = 1.47, z = -3.44, p < .001). There were no differences between the two HWD conditions regarding sound experience, device use, or display use.

Discussion

The data show different patterns in session 1 versus session 2 and 3, but they provide some support for our hypotheses. In session 1, participants answered the SAGAT questions more accurately in the HWD-notifications condition than in either the alarms only condition or the HWD-alarms condition. These findings partially support our hypotheses H1 and H2, and suggest that the combination of the more informative notifications and the HWD display improved participants' ability to maintain continuous awareness of patient related problems rather than the visual display by itself. After session 1, however, situation awareness in the HWD-alarms and alarms only conditions rose to the same level as the HWD-notifications condition.

In session 1, the alarms in the HWD-alarms condition may not have adequately cued the participant's visual sampling behavior. The notifications in the HWD-notifications condition, however, were tied to specific changes, and could be used to guide visual behavior. The notifications may have contributed to an increase in accuracy, but the SAGAT questions were specific enough, and the notifications ambiguous enough, that participants probably still needed the HWD display to answer the questions accurately.

Performance on the patient assessment task with vSim task was significantly better in both HWD conditions compared to the alarms only condition, supporting our third hypothesis, H3. The HWD conditions probably supported more accurate vSim performance because participants did not have to stop their assessment and move to the patient "ward" to get a patient update.

Participants also reported a significant reduction in workload while using the HWD-alarms and HWD-notifications conditions compared with the alarms only condition, supporting our fourth hypothesis, H4. Participants in the HWD conditions could instantly check the status of a patient without having to interrupt their patient assessment task to physically move to another patient's room.

Finally, participants reported that the notifications in the HWD-notifications condition were less annoying than the alarms in the alarms only condition, and that the alarms in the HWD-alarms condition were less annoying than the alarms in the alarms only condition. We expected in hypothesis H5, that the alarms in both conditions would be reported as more annoying than the notifications, so this result was unexpected. Taken with the reduction in workload, it would seem that the HWD generally improved the experience of the experiment which included their opinions of the alarms in the HWD-alarms condition.

General discussion

Experiments 1 and 2 confirmed our hypothesis that HWDs would offer benefits in a simulated multiple patient monitoring context. In Experiment 1, participants using the HWD performed better on the patient monitoring task by managing more clinical alarms than participants without the continuous display, and by managing them faster. These results occurred regardless of whether there was a low or high frequency of alarms, and without affecting performance on an ongoing cognitive task. In Experiment 2, participants using the HWD answered situation awareness questions more accurately, but only when the HWD was paired with a more informative auditory display than standard bedside alarms, and only in the first session. In subsequent sessions, there was no apparent difference in situation awareness across the displays. However, in both HWD conditions of Experiment 2, participants performed better on the patient assessment task, reported less workload, and reported that the auditory displays were less annoying. Together our findings suggest that there may be benefits of using HWDs in multiple patient contexts. These findings extend the potential advantages for HWDs beyond previously-demonstrated benefits for anesthetists during surgery (Liu, Jenkins, & Sanderson, 2009a; Liu, Jenkins, et al., 2009b; Vorraber et al., 2014).

One benefit afforded by the HWD is the ability to disambiguate alarms. In Experiment 1, participants with the continuous information on the HWD not only treated their patients faster, but also delayed their responses to clinically irrelevant alarms, therefore spending more time performing the ongoing task, compared to participants without the HWD. In a clinical setting, a doctor or nurse wearing an HWD and facing multitasking demands could make better tactical decisions about the allocation of their attention. Similarly, a team leader wearing the HWD would know which patient or patients had become unstable, and would be ready to offer assistance to staff. Finally, the HWD would also reassure clinicians when all patients' vital signs are stable, which otherwise would not be possible without continual visits to the patient's bedside or to a centralized monitoring station.

A second benefit of the HWD is participants' ability to maintain awareness of patients' status. As noted above, in Experiment 2, the HWD was associated with better situation awareness, but only in the first session and only when coupled with the auditory notifications. We had expected the benefits of mobile access to continuous information to persist across sessions, and we expected that even the HWD-alarms condition would improve situation awareness over the alarms only condition. A potential explanation for the initial benefit of the notifications over alarms in the HWD conditions is that the timing of the notifications was more

consistent with the actual changes in a patient's state. The alarm sounds repeated every 15 seconds, so it was more difficult to distinguish an old alarm with a new alarm. Moreover, the notifications conveyed specific information from each patient potentially allowing the participant to know which patient changed, even they couldn't specifically determine what that change was. The improvement in situation awareness for HWD-alarms and alarms only in session 2 and session 3 may have arisen from carryover effects, after participants had experienced at least one HWD condition. In addition, the improvement may have arisen from similarities between scenarios and questions, which were all matched and limited in complexity. The simplicity of the vital signs (only three per patient) and limited number of changes that could occur may have made the SAGAT questions more predictable over time, removing differences between display conditions.

Importantly, participants using the HWD achieved better performance despite the potential for certain challenges that often accompany the use of HWD, as outlined in the Introduction. In most studies where such challenges are evident, participants were asked to keep their visual attention focused on both the outside world and the HWD. In the present studies, however, the information on the continuous display was viewed at the participant's discretion, such as when they wanted an update on a simulated patient's status or to check the nature of an alarm or notification. Therefore the HWD was used in a way that would have overcome the above challenges.

Limitations and future research

A first limitation is that the results are not sufficient to conclude that continuous information displays, and specifically HWDs, will support clinicians adequately in healthcare settings. Although many of the stimuli and tasks drew from stimuli and tasks in a clinical environment, the experiments were still very simplified. As an example, performance on SAGAT measures in nursing often falls below 50% (Bogossian et al., 2014). In the current study however, participants' SAGAT performance ranged from 50% to 70%, even without advanced displays. In addition, the alarm types (clinical vs. sensor) were simplified in that they were easy to discriminate. Further tests should use fully representative patient data to more accurately portray both actionable, and nonactionable alarms.

A second potential limitation arises from the ongoing tasks used in each experiment. Neither task could simulate the high levels of stress, and the necessity for precision, that accompanies work in a hospital environment. A more representative ongoing task, and one that that imposed costs if not performed well, may affect participants' ability to use the HWD

differently, as suggested in a growing body of literature (Liu, Jenkins, Sanderson, et al., 2009; Pascale et al., 2015; Ververs & Wickens, 1998; Winterbottom, Patterson, Pierce, Gaska, & Hadley, 2015; Woodham et al., 2016).

Nevertheless, an HWD does not necessarily need to be continuously monitored to be effective. An HWD can be supplemented with auditory alerts that capture attention when needed, rather than requiring the wearer to rely solely on the visual display to capture visual attention (Barde, Lee, Ward, Helton, & Billinghurst, 2016; Browning, 2016). For example, the notification display used in Experiment 2 offered an effective, albeit simplified, means of cueing the participants' HWD sampling behavior, but there are undoubtedly better means (Hickling, Brecknell, Loeb, & Sanderson, 2017; Li et al., 2017). Furthermore, with well-designed visual displays (Koch et al., 2013; Zhang et al., 2002), a clinician may be able to glance rapidly to check for patient-related concerns while maintaining an acceptable level of performance on the task at hand.

Conclusion

The purpose of the current series of studies was to test whether continuous access to simulated patient information, presented via an HWD, could improve alarm management behavior and situation awareness in a controlled laboratory test with a simulated hospital microworld. The HWD supported better alarm prioritization practices, and an increase in the amount of time participants could spend performing an ongoing task. Participants' situation awareness was enhanced by the HWD when data from their first session were compared, and only when the HWD was coupled with an informative notification display. Under more constrained conditions, however, when physically stopping an ongoing task is not possible, the HWD with its continuous information may outperform other displays. If these effects were to translate to a clinical environment, clinicians may be able to use similar displays to acknowledge the source and urgency of an alarm, to facilitate more appropriate responses, and to maintain greater awareness of the statuses of multiple patients simultaneously without having to interrupt the task at hand.

Key Points

- Hospital clinicians are mobile workers who often move away from their patients' bedsides or from central patient monitoring displays.
- We showed that continuous monitoring of simulated patients with a head-worn display (HWD) helped participants prioritize patients for care and increased awareness for the status of multiple patients.
- Ongoing task performance was not compromised when participants used an HWD, and instead tended to improve.
- Further research is necessary with participants and tasks that are more representative of hospital monitoring environments.

Acknowledgements

This research was supported by Australian Research Council Discovery Project DP140101822 to Sanderson, Loeb, and Liu. Liu is supported by a Queensland Health Junior Doctor Research Fellowship. We thank members of the Cognitive Engineering Research Group for contributing valuable feedback.

References

Baig, M. M., GholamHosseini, H., & Lindén, M. (2015). *Tablet-based patient monitoring and decision support systems in hospital care.* Paper presented at the Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE.

Barde, A., Lee, G., Ward, M., Helton, W. S., & Billinghurst, M. (2016). *A bone conductionbased spatial auditory display as part of a wearable hybrid interface*. Paper presented at the The 22nd International Conference on Auditory Display, Canberra, Australia.

Bogossian, F., Cooper, S., Cant, R., Beauchamp, A., Porter, J., Kain, V., . . . Phillips, N. M. (2014). Undergraduate nursing students' performance in recognising and responding to sudden patient deterioration in high psychological fidelity simulated environments: An Australian multi-centre study. *Nurse Education Today, 34*(5), 691-696. doi:http://dx.doi.org/10.1016/j.nedt.2013.09.015

Browning, C. (2016). *Head Worn Displays with Visual Cues and Animal Call Alerts for Multiple Patient Monitoring.* (BPsySci(Hons)), The University of Queensland, St Lucia, Australia.

Cvach, M. M., Frank, R. J., Doyle, P., & Stevens, Z. K. (2014). Use of pagers with an alarm escalation system to reduce cardiac monitor alarm signals. *Journal of Nursing Care Quality, 29*(1), 9-18.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32-64.

Gorges, M., Markewitz, B. A., & Westenskow, D. R. (2009). Improving alarm performance in the medical intensive care unit using delays and clinical context. *Anesthesia and Analgesia, 108*(5), 1546-1552. doi:10.1213/ane.0b013e31819bdfbb

Graham, K. C., & Cvach, M. (2010). Monitor alarm fatigue: Standardizing use of physiological monitors and decreasing nuisance alarms. *American Journal of Critical Care, 19*(1), 28-34. doi:10.4037/ajcc2010651

He, J., Ellis, J., Choi, W., & Wang, P. (2015). *Driving while reading using Google glass versus using a smart phone: which is more distracting to driving performance?* Paper presented at the Proceedings of the Eighth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.

Hickling, A., Brecknell, B., Loeb, R. G., & Sanderson, P. (2017). Using a sequence of earcons to monitor multiple simulated patients. *Human Factors, 59*(2), 268-288.

IEC60601-1-8, I. I. S. (2005– 08). Medical electrical equipment, Part 1– 8: general requirements for safety— collateral standard: general requirements, tests and guidance for alarm systems in medical electrical equipment and medical electrical systems. Geneva, Switzerland: International Electrotechnical Commission.

Imhoff, M., Kuhls, S., Gather, U., & Fried, R. (2009). Smart alarms from medical devices in the OR and ICU. *Best Practice & Research: Clinical Anaesthesiology, 23*(1), 39-50. doi:10.1016/j.bpa.2008.07.008

Koch, S. H., Weir, C., Westenskow, D., Gondan, M., Agutter, J., Haar, M., . . . Staggers, N. (2013). Evaluation of the effect of information integration in displays for ICU nurses on situation awareness and task completion time: a prospective randomized controlled study. *International Journal of Medical Informatics, 82*(8), 665-675.

Lavie, N. (2005). Distracted and confused?: selective attention under load. *Trends Cogn Sci, 9*(2), 75-82. doi:10.1016/j.tics.2004.12.004

Lavie, N., Beck, D. M., & Konstantinou, N. (2014). Blinded by the load: attention, awareness and the role of perceptual load. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 369*(1641), 20130205. doi:10.1098/rstb.2013.0205

Li, S. Y., Tang, T.-L., Hickling, A., Yau, S., Brecknell, B., & Sanderson, P. M. (2017). Spearcons for Patient Monitoring: Laboratory Investigation Comparing Earcons and Spearcons. *Human Factors*, 0018720817697536.

Liu, D., Jenkins, S. A., & Sanderson, P. M. (2009a). *Clinical implementation of a head-mounted display of patient vital signs.* Paper presented at the 2009 International Symposium on Wearable Computers, Proceedings.

Liu, D., Jenkins, S. A., & Sanderson, P. M. (2009b). Patient monitoring with head-mounted displays. *Current Opinions in Anaesthesiology, 22*(6), 796-803. doi:10.1097/ACO.0b013e32833269c1

Liu, D., Jenkins, S. A., Sanderson, P. M., Watson, M. O., Leane, T., Kruys, A., & Russell, W. J. (2009). Monitoring with head-mounted displays: performance and safety in a full-scale simulator and part-task trainer. *Anesthesia and Analgesia, 109*(4), 1135-1146. doi:10.1213/ANE.0b013e3181b5a200

Mustonen, T., Berg, M., Kaistinen, J., Kawai, T., & Hakkinen, J. (2013). Visual task performance using a monocular see-through head-mounted display (HMD) while walking. *Journal of Experimental Psychology: Applied, 19*(4), 333-344. doi:10.1037/a0034635

Paine, C. W., Goel, V. V., Ely, E., Stave, C. D., Stemler, S., Zander, M., & Bonafide, C. P. (2016). Systematic Review of Physiologic Monitor Alarm Characteristics and Pragmatic Interventions to Reduce Alarm Frequency. *Journal of Hospital Medicine, 11*(2), 136-144. doi:10.1002/jhm.2520

Pascale, M., Sanderson, P., Liu, D., Mohamed, I., Stigter, N., & Loeb, R. (2015). *Peripheral detection for abrupt onset stimuli presented via head-worn display.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Preece, M., Horswill, M., Hill, A., & Watson, M. (2010). The development of the Adult Deterioration Detection System (ADDS) chart. Sydney. *New South Wales, Australia: Australian Commission on Safety and Quality in Health Care*.

Raveh, D., & Lavie, N. (2015). Load-induced inattentional deafness. *Attention, Perception, & Psychophysics, 77*(2), 483-492.

Royal Prince Alfred Hospital Patient Observation (Vital Signs) Policy – Adult. (2010). Retrieved from Sydney, NSW, Australia: https://www.safetyandquality.gov.au/wp-content/uploads/2012/02/RPA-observations-policy-directive.pdf

Ruskin, K. J., & Hueske-Kraus, D. (2015). Alarm fatigue: impacts on patient safety. *Current Opinion in Anesthesiology, 28*(6), 685-690.

Sanderson, P. M. (2006). The multimodal world of medical monitoring displays. *Applied Ergonomics, 37*(4), 501-512. doi:10.1016/j.apergo.2006.04.022

Sawyer, B. D., Finomore, V. S., Calvo, A. A., & Hancock, P. A. (2014). Google glass: A driver distraction cause or cure? *Human Factors, 56*(7), 1307-1321. doi:10.1177/0018720814555723

Stubbings, L., Chaboyer, W., & McMurray, A. (2012). Nurses' use of situation awareness in decision-making: an integrative review. *Journal of Advanced Nursing, 68*(7), 1443-1453.

Ververs, P. M., & Wickens, C. D. (1998). Head-up displays: Effect of clutter, display intensity, and display location on pilot performance. *The International Journal of Aviation Psychology, 8*(4), 377-403.

Vorraber, W., Voessner, S., Stark, G., Neubacher, D., DeMello, S., & Bair, A. (2014). Medical applications of near-eye display devices: an exploratory study. *International Journal of Surgery (London, England), 12*(12), 1266-1272. doi:10.1016/j.ijsu.2014.09.014

Watkins, T., Whisman, L., & Booker, P. (2016). Nursing assessment of continuous vital sign surveillance to improve patient safety on the medical/surgical unit. *Journal of Clinical Nursing, 25*(1-2), 278-281. doi:10.1111/jocn.13102

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. doi:DOI 10.1518/hfes.45.3.360.27250

Wickens, C. D., & Long, J. (1995). Object Versus Space-Based Models of Visual-Attention - Implications for the Design of Head-up Displays. *Journal of Experimental Psychology-Applied, 1*(3), 179-193. doi:Doi 10.1037//1076-898x.1.3.179

Winterbottom, M. D., Patterson, R., Pierce, B. J., Gaska, J., & Hadley, S. (2015). *Visibility of monocular symbology in transparent head-mounted display applications.* Paper presented at the SPIE Defense+ Security.

Woodham, A., Billinghurst, M., & Helton, W. S. (2016). Climbing With a Head-Mounted Display: Dual-Task Costs. *Human Factors, 58*(3), 452-461. doi:10.1177/0018720815623431

Xiao, Y., Seagull, F. J., Nieves-Khouw, F., Barczak, N., & Perkins, S. (2004). Organizational-historical analysis of the" failure to respond to alarm" problems. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 34*(6), 772-778.

Zhang, Y., Drews, F., Westenskow, D. R., Foresti, S., Agutter, J., Bermudez, J. C., . . . Loeb, R. (2002). Effects of integrated graphical displays on situation awareness in anaesthesiology. *Cognition, Technology & Work, 4*(2), 82-90.

Appendix A

Alarm ranges in Experiment 1

Table A1

Numerical ranges for each vital sign in Experiment 1.

Levels	Heart Rate (HR)	Blood Pressure (BP)	Oxygen Saturation (SpO2)
High priority alarm state (high)	120—130	150—155	Not applicable
Trending upwards	106—119	136—149	Not applicable
Normal	75—105	95—135	97—100
Trending downwards	61—74	81—94	90—96
High priority alarm state (low)	50—60	75—80	85—89

The normal and abnormal ranges used for each vital sign in Experiment 1 were derived from the Royal Prince Alfred Hospital Patient Observation Policy (*Royal Prince Alfred Hospital Patient Observation (Vital Signs) Policy – Adult*, 2010) and are shown in Table A1. Alarms were triggered when the value of the vital sign crossed into the high priority range. The alarm sounded every 15 seconds until the alarm crossed back outside of the high priority range. Alarm sounds were accompanied by color inversion on the visual displays of the vital signs, either at the microworld, while inside a patient's room, or on the HWD.

Alarm ranges in Experiment 2

Table A2

Numerical ranges for each vital sign in Experiment 2.

Levels	Heart Rate (HR)	Blood Pressure (BP)	Oxygen Saturation (SpO2)
Critical - high	≥ 130	≥ 200	Not applicable
Warning - high	100 - 129	160 – 199	Not applicable
Normal	50-99	110 – 159	95-100
Warning - low	40-49	90 – 109	85-94
Critical - low	≤ 39	≤ 89	≤ 84

For Experiment 2, changes occurred when one or more vital signs crossed a predetermined threshold given in the Queensland Adult Deterioration Detection System (Q-ADDS) (Preece, Horswill, Hill, & Watson, 2010) (see Table A2). In Experiment 2, we included two alarm types:

one for warning alarms (in the warning range, and one for critical alarms (in the critical range). The alarms sounded anytime a patient's status changed outside of the normal range, and then every 15 seconds while that patient state persisted. There was only one alarm sound for each patient, regardless of how many vital signs were outside the normal range for that patient. So, at any given time, there could be up to six alarms, each with its own 15 second interval.

The notification display sounded every time a patient's status changed, like the alarms, but also when a patient returned to a normal state. If a patient's status did not change, the notification display was set to sound after 30 seconds. Each instance of the notification display played a sound for all six patients, in order, so that participants could potentially keep track of who's status was changing without having to look at the visual display. For any further information, like which vital sign and which direction, the participant had to look at the visual display.

Author biographies

Michael Pascale is a PhD candidate at The University of Queensland. He completed his MA in Psychology at the University of North Carolina at Wilmington in 2013.

Penelope Sanderson is Professor of Cognitive Engineering and Human Factors at The University of Queensland, with appointments in the Schools of Psychology, of Information Technology and Electrical Engineering, and of Clinical Medicine. She received her PhD from University of Toronto in 1985.

David Liu is a Postdoctoral Research Fellow in the Faculty of Medicine at The University of Queensland, where he received his PhD in 2010 and his MBBS in 2015. He was a Fulbright scholar at the University of Utah in 2009.

Ismail Mohamed is a research engineer in the Cognitive Engineering Research Group at The University of Queensland. He received his BEng(Hons) in 2006 from The University of Queensland. His research interests include signal processing, application development and digital communication.

Birgit Brecknell is a research scientist with the Cognitive Engineering Research Group at the University of Queensland. She received a BEng(Hons) in aerospace engineering in 2001, an MA in screenwriting in 2008, and a PhD in computer vision in 2005, all from Queensland University of Technology.

Robert G. Loeb is a clinical professor of anesthesiology at University of Florida–Gainesville and an honorary associate professor of psychology at The University of Queensland. He received his MD in 1983 from University of Maryland.

Appendix C: Ethics Approval – Study 1

Friday, 3 October 2014 1:48:26 PM Australian Eastern Standard Time

Subject: Ethics Application
Date: Friday, 3 October 2014 1:44:17 PM Australian Eastern Standard Time
From: Jeanie Sheffield
To: Michael Pascale
CC: Penelope Sanderson, Danico Jones

Dear Michael

Many thanks for submitting your application for ethical approval of your student research project: **Task performance with a simulated head-worn display**

I have now reviewed your application and am happy to provide approval for the study.

The clearance number for the study is: **14-PSYCH-PHD-59-JS**

Thanks for submitting such a thorough and clear application – it makes my job much easier.

I think the research sounds really interesting and I hope you will find some good results.

All the best

Jeanie

Jeanie Sheffield PhD

Room 122 School of Psychology
The University of Queensland
Brisbane Queensland Australia 4072
Phone: + 61 7 3365 6690
Fax: + 61 7 3365 4466
Email: jeanie@psy.uq.edu.au

Appendix D: Ethics Amendment Approval – Study 2

Tuesday, 13 January 2015 6:00:22 PM Australian Eastern Standard Time

Subject: Re: Michael Pascale - Ethics Amendment
Date: Tuesday, 13 January 2015 5:57:48 PM Australian Eastern Standard Time
From: Jeanie Sheffield
To: Psychology Ethics, Michael Pascale, Penelope Sanderson

Dear Michael

Thanks for submitting the amendment to your ethics application. I am happy to grant approval for the amendments. I'm sure you will do this and I may have missed it in the attachments but it would be good to ensure that participants feel comfortable wearing the equipment and that they can let you know if they are not

The approval number will remain the same

Regards

Jeanie

Sent from my iPad

Appendix E: Ethics Amendment Approval – Study 3

Subject: Ethics Amendment
From: Jeanie Sheffield <j.sheffield@psy.uq.edu.au>
Date: 21/10/15 11:04
To: Mike Pascale <m.pascale@uq.edu.au>
CC: Psychology Ethics <psyethics@psy.uq.edu.au>, Penelope Sanderson <psanderson@itee.uq.edu.au>

Dear Mike

Thanks for submitting a further amendment to your original ethics application [14-PSYCH-PHD-59-JS].

I am happy to provide approval for the amendment subject to one minor change:

On both your participant information sheet and consent form you will need to remove Jolanda Jetten as one of the school ethics officer and replace her name with Julie Henry. If you have a look at the current ethics application form, Julies's contact details are shown on the Participant Information Sheet template. You do not need to send in any further materials.

All the best with this new addition to your research - it sounds more realistic for participants. Let's hope they can keep some 'patients' alive

Regards

Jeanie

Appendix F: Ethics Amendment Approval – Study 4

Subject: Ethics Amendment
From: "Jeanie Sheffield" <j.sheffield@psy.uq.edu.au>
Sent: 26-Apr-16 9:08:28 AM
To: "Mike Pascale" <m.pascale@uq.edu.au>
CC: "Penelope Sanderson" <psanderson@itee.uq.edu.au>;"Psychology Ethics" <psyethics@psy.uq.edu.au>

Dear Michael

I have read your application for amendment to your research study: **Task performance with a simulated head-worn display**

I am happy to approve those amendments. Your approval number will remain the same.

Regards

Jeanie

Jeanie Sheffield PhD

Room 122 School of Psychology
The University of Queensland
Brisbane Queensland Australia 4072
Phone: + 61 7 3365 6690
Fax: + 61 7 3365 4466
Email: jeanie@psy.uq.edu.au

Appendix G: Ethics Amendment Approval – Study 5

Subject: Ethics Amendment
From: "Jeanie Sheffield" <j.sheffield@psy.uq.edu.au>
Sent: 25-Sep-17 14:16:46
To: "Mike Pascale" <m.pascale@uq.edu.au>;
CC: "Penelope Sanderson" <psanderson@itee.uq.edu.au>; "Psychology Ethics" <psyethics@psy.uq.edu.au>;

Dear Michael

Thanks for submitting your request for an amendment to your research project: **Task performance with a simulated head-worn display**

I have reviewed the application and am happy to provide approval for you to:

- Extend the opportunity to participate to The University of Southern Queensland and Southern Cross University (gatekeeper letters of approval from them attached with your amendment request).

The approval number remains the same: 14-PSYCH-PHD-59-JS

All the best with the study and I hope the recruitment is successful

Regards

Jeanie Sheffield PhD

Room 122 School of Psychology
The University of Queensland
Brisbane Queensland Australia 4072
Phone: + 61 7 3365 6690
Fax: + 61 7 3365 4466
Email: jeanie@psy.uq.edu.au