Informative Vibrotactile Displays to Support Attention and Task Management in Anesthesiology

by

Thomas Keith Ferris

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Operations Engineering)
in The University of Michigan
2010

Doctoral Committee:

Associate Professor Nadine B. Sarter, Chair
Professor Yili Liu
Associate Professor Brent Gillespie
Associate Professor Bernard J. Martin
Clinical Lecturer Robert Edwin Christensen

Acknowledgments

I would like to acknowledge the following people and sources of support that have made this dissertation possible. First, I thank the National Science Foundation (NSF), Rackham Graduate School, and the Department of Industrial and Operations Engineering (IOE) for the financial support each provided to me during my graduate studies at the University of Michigan. Additionally, I'd like to acknowledge NSF, Army Research Laboratory, and Air Force Research Laboratory for grants which supported much of the research described in this dissertation.

I thank my advisor and mentor, Nadine Sarter, for 6+ years of support and encouragement. I consider myself very fortunate to have had the opportunity to train and work alongside such an excellent teacher and researcher, and am very grateful for her efforts during my time at Michigan and for the future support I'm sure she will provide. The confidence I have in taking the next steps of my academic career is a testament to Nadine's ability to consistently challenge me as a researcher, and to provide just the right encouragement when the challenge seemed overwhelming.

I'd also like to thank my other dissertation committee members: Yili Liu, Bernard Martin, Brent Gillespie, and Robert Christensen, for their time, support, and valuable input during the dissertation process. I extend this thanks to the faculty and staff of the IOE department and the Center for Ergonomics, who have helped me in numerous ways throughout my graduate career. The Society for Technology in Anesthesia and the UM Department of Anesthesiology also deserve my gratitude, for helping an engineering student learn about the fascinating field of anesthesiology, and for giving me access to the hospital environment and so much of their valuable time. In particular, I'd like to thank Kevin Tremper, Michael O'Reilly, James Szocik, Amy Shanks, and Michelle Morris for their help in various stages of my research over the past 5 years, and I'd again like to recognize Robert Christensen for his help in this regard. Without his enthusiasm for my project and his willingness to meet with me and address endless questions about anesthesiology, I could not have conducted meaningful research in this field.

I'd like to thank the THInC lab, in particular the members who directly contributed to the work described in this dissertation: Shameem Hameed, Joe Phillips, Clif Reeder, Hari Molabanti, Kiana Stringfield, Lantei Takona, and Nikhil Rao. All members of the lab, as well as my other IOE classmates, deserve my thanks for providing insightful input to (and participating in) my research projects, and for their friendship and emotional support.

I'd like to thank my family and friends, many of whom have already been mentioned, for helping to keep me mentally, emotionally, and spiritually grounded during my graduate studies. I especially want to thank my parents Michael and Susan, and my parents-in-law Paul and Brenda, for all the advice, encouragement, and love they have provided.

Finally and most importantly, I thank my wife Erin. During my graduate studies, she has been my greatest supporter, celebrating my successes with me and encouraging me while weathering the storms of stress and frustration. She deserves as much credit as I do for the accomplishment of this dissertation. She is a wonderful partner, mother, and friend, and I'm lucky to have her with me as we begin the next phase of our lives.

Table of Contents

Acknowledgments ..... **ii**
List of Figures ..... **vi**
List of Tables ..... **ix**
List of Equations ..... **x**
List of Appendices ..... **xi**
Chapter 1: Introduction **1**
The “alarm problem” in patient monitoring 3
From alarms to improvements in continuous informing 7
The potential for tactile displays to support patient monitoring 13
The case for a “tactification” physiological data display system 17
References 26
Chapter 2: Two Important Considerations in Tactile Interface Design: Tactile Change Blindness and Processing Code Interference **33**
Part A: Tactile “Change Blindness” in the Detection of Vibration Intensity **35**
Introduction 35
Method 36
Results 42
Discussion 45
References 48
Part B: The Role of Processing Code in Tactile Display Design **50**
Introduction 50
Method 56
Results 64
Discussion 71
References 75
Contributions to Tactification Design **78**
Chapter 3: Survey of Anesthesia Providers **80**
Survey Response Summary 83
Chapter 4: A First Investigation of the Use of Vibrotactile Displays to Support Physiological Monitoring and Multitasking: Iconic Tactile Displays **92**
Introduction 93
Method 94

Results 99
Discussion 101
References 104
Chapter 5: Design and Evaluation of Tactification Displays 105
Tactile Displays 106
Design Considerations 109
Tactile Alarm and Tactification Display Designs 118
Hypotheses 125
Evaluation Study 132
Results and Discussion 163
General Discussion 201
References 209
Chapter 6: Conclusion 212
References 219
Appendices 221

List of Figures

Figure 2.1: The simulation setup, distributed across three separate desktop monitors. 37
Figure 2.2: Intensity changes for the five presentation conditions 39
Figure 2.3: Hit rate for each presentation condition and each magnitude of change in the isolation and dual-task cases 43
Figure 2.4: SDT analysis results. 44
Figure 2.5: Adapted model of the structure of processing resources for a visual/tactile task set.... 53
Figure 2.6: Schematic of an “obstacle zone” in the driving scenario.. 58
Figure 2.7: Example images displayed for each of the two visual tasks. 59
Figure 2.8: Example of the vibration sequence for a spatial tactile icon, and the required button-press response. 61
Figure 2.9: Four pulse patterns which could be displayed for a vibration segment of a nonspatial tactile icon. 62
Figure 2.10: Response accuracies (a) and times (b) for the identification of spatially-encoded tactile icons in the single-task T condition and the dual-task TS and TN conditions in Experiment 1 65
Figure 2.11: Visual task accuracies (% correct lane choice) for the visual-spatial and visual-nonspatial tasks in each respective single- and dual-task condition in Experiment 1 (spatially-encoded icons). 66
Figure 2.12: Response accuracies (a) and times (b) for the identification of nonspatially-encoded tactile icons in the single-task *T* condition and the dual-task *TS* and *TN* conditions in Experiment 2. 68
Figure 2.13: Visual task accuracies (% correct lane choice) for the visual-spatial and visual-nonspatial task in each respective single- and dual-task condition in Experiment 2 (nonspatially-encoded icons). 69
Figure 3.1: Screenshot of the online survey. 82
Figure 3.2: Responses to questions about how much conducting each of the above tasks interferes with the ability to monitor blood pressure. 84
Figure 3.3: “Usefulness” scores for various types of information, relating how useful each is for the task of blood pressure management under routine conditions. 85
Figure 3.4: Box plots of responders’ ratings of “obstructiveness” of a vibrotactile display at several body sites. 87

Figure 3.5: Box plots of responders' ratings of "annoyance" for vibration presentations to several body sites. 87
Figure 3.6: Box plots of responders' ratings of "distraction" that would result from vibration presentations at several body sites. 88
Figure 3.7: Response frequency for vibration dimensions/modulations that would most intuitively communicate each of the five most useful types of blood pressure information. 91
Figure 4.1: The simulation setup, distributed across three separate desktop monitors. 95
Figure 4.2: Desktop display of physiological parameters 96
Figure 4.3: Results related to MAP management 101
Figure 5.1: The C-2 tactor developed by Engineering Acoustics, Inc. 107
Figure 5.2: Male compression garment used to secure tactors to the back and upper arm (Design Veronique) 108
Figure 5.3: Experimental garment with tactors attached. 109
Figure 5.4: The spatial arrangement of the TV display was designed to feel like two lungs "filling" with each breath. 111
Figure 5.5: The metaphorically-derived pattern used to relate ETCO2 levels was designed to feel like the capnographic waveform. 113
Figure 5.6: Signal adaptations with respect to changing health state. 119
Figure 5.7: An example of the "beat pulse" signal activation pattern. 120
Figure 5.8: The "fill" pattern representing normal TV levels 122
Figure 5.9: Fill patterns associated with "trending" low TV levels (left) and "problem" low TV levels (right) 123
Figure 5.10: Fill patterns associated with "trending" high (left) and "problem" high (right) TV levels. 123
Figure 5.11: Layout of visual displays for the TAPIS simulator. 134
Figure 5.12: Participant field of view when facing the patient display (center). 136
Figure 5.13: View to the right of the participant (when turned away from the patient display). 137
Figure 5.14: Views of the two tool tables, accessed by "turning" to the left or right of the patient, by way of moving the mouse. 138
Figure 5.15: Induction task steps depicting the insertion of the endotracheal tube, which involved using the tethered wiimote to control the laryngoscope and the mouse to insert the tube. 139
Figure 5.16: Physiological display 1, highlighting the parameter MAP, which was one of the three critical parameters for the physiological monitoring task. 140
Figure 5.17: Physiological display 2, highlighting the parameters ETCO2 and TV, two of the three critical parameters for the physiological monitoring task. 141
Figure 5.18: Infusion pump display (touchscreen interface). 144
Figure 5.19: The Troubleshooting menu, which was displayed whenever the BRB was pressed. 146
Figure 5.20: Calculating physiological management score. 158
Figure 5.21: Detection and Correction times for all events, by display configuration. 168
Figure 5.22: Detection and Correction times for all events, by scenario. 169

Figure 5.23: a) Detection times and b) Correction times for events in Scenario 1...................... 170
Figure 5.24: a) Detection times and b) Correction times for events in Scenario 2...................... 171
Figure 5.25: a) Detection times and b) Correction times for events in Scenario 3...................... 172
Figure 5.26: a) Detection times and b) Correction times for events in Scenario 4...................... 173
Figure 5.27: a) Detection times and b) Correction times for blood pressure and respiration events, by speed of event development. 175
Figure 5.28: a) Detection times and b) Correction times for blood pressure and respiration events, by demand during event occurrence 177
Figure 5.29: Physiological management score according to display configuration for a) MAP, b) ETCO2, and c) TV. 184
Figure 5.30: Induction completion times under each display configuration. 188
Figure 5.31: Induction completion times for each scenario. 188
Figure 5.32: Mean multitask performance scores, calculated with Equation 5.1, for each display configuration. 191
Figure 5.33: Mean ratings for each display configuration (key in upper right corner of (b)) for the six attributes on the NASA-TLX survey: a) Mental demand, b) Physical demand, c) Temporal demand, d) Performance, e) Effort, and f) Frustration... 194
Figure 5.34: Box plots of subjective rankings of each display configuration, according to the attributes listed on the postexperiment questionnaire (see Appendix 8) 196

List of Tables

Table 2.1: Summary of task condition notations 64
Table 2.2: Dual-task performance metrics for each combination of tactile icon and visual task.. 71
Table 5.1: Target values and "acceptable" ranges for each of the three critical parameters in the physiological monitoring task 142
Table 5.2: Serious health events, listed in order of their occurrence within the induction task, summary descriptions of the events, and relative properties (continued on next three pages)...... 148
Table 5.3: Balanced design used to assign display configurations to scenarios for each participant. 154
Table 5.4: Classification of blood pressure and respiration events, according to the independent variables speed of event development and (visual/attentional) demand during event occurrence 156
Table 5.5: Detection rates (%) for each participant by a) Scenario and b) Display configuration..... 164
Table 5.6: Detection rates in each display configuration for each individual event, and events broken down by independent measures of type (blood pressure, respiration, emergency), speed of development (fast- or slow-developing), and demand during event occurrence (high or low) 165
Table 5.7: Significance levels and mean subjective rank order of display configurations for each of the attributes listed on the postexperiment questionnaire (see Appendix 8). 198

List of Equations

Equation 2.1: Metric for dual-task performance in conditions which paired visual-task *V* (Spatial or Nonspatial) and tactile icons defined by encoding method *t* (spatial or nonspatial) 71
Equation 5.1: Multitask combined performance score (P_{DS}). 159

List of Appendices

Appendix 1: Complete List of Online Survey Questions 221
Appendix 2: TAPIS Patient Chart 234
Appendix 3: TAPIS Induction Task Steps 235
Appendix 4: Normal Fluctuation Script Example 239
Appendix 5: Example Scripts for Blood Pressure, Respiration, and Emergency Events 242
Blood Pressure Event Script Example: Hyper_nervous 242
Respiration Event Script Example: Hyperventilation 243
Emergency Event Script: Anaphylaxis 244
Appendix 6: Hybrid Display "Pre-test" Presentations and Participant Scores 250
Appendix 7: NASA-TLX Subjective Ratings Worksheet 253
Appendix 8: Postexperiment Questionnaire 254

Chapter 1

Introduction

Preventable medical error has been one of the leading causes of death in the United States during the past several decades (Leape, 1994; Kohn, Corrigan, & Donaldson, 2000; Healthgrades, 2008) and remains a major concern today. A recent study reported that between 2004 and 2006 in the U.S. alone, 238,000 deaths and $8.8 billion in healthcare-related costs could be directly attributed to errors on the part of healthcare providers (Healthgrades, 2008). To put the death toll in perspective, the average American is more than twice as likely to die from a preventable medical error as from a motor vehicle accident (Kohn et al., 2000). According to the landmark Institute of Medicine report, "To Err is Human," the most critical factor in patient safety and the occurrence of these preventable errors is breakdowns in human attention (Kohn et al., 2000).

Anesthesiologists face one of the most attentionally-demanding task sets in the clinical setting, especially in regard to the ability to divide attention between multiple tasks and sources of task-relevant data. The practice of anesthesiology involves monitoring a large number of life-critical patient health parameters, including the patient's oxygenation (O_2 delivery and hemoglobin saturation), ventilation (breathing rate and volume), circulation (heart rate and blood pressure), core body temperature, and other operation-specific parameters (ASA, 1986; 2005). The monitoring task is conducted in parallel with other demanding tasks related to patient care, such as preparing and administering various intravenous drugs, programming drug infusion machines, and preparing the anesthetized patient for mechanical ventilation (e.g., clearing the airway and installing a rigid tube in the trachea). Errors in the monitoring task (i.e., “monitoring errors”), which are often the result of inefficient allocation of attention to the spatially distributed sources of patient physiological data, constitute a significant portion of anesthesia-related patient injuries (Cooper, Newbower, Long, & McPeek, 1978; Walsh & Beatty, 2002; Webb et al., 1993). These errors tend to be most prevalent when the ability to effectively distribute attention among the sources breaks down because the anesthesiologist is interrupted (e.g., Grundgeiger, Liu, Sanderson, Jenkins, & Leane, 2008) or distracted by an usual task (e.g., Sanderson et al., 2008), and during the induction and emergence phases of anesthesia, which are characterized by a higher cognitive workload and demand for multitasking (Gaba & Lee, 1990; Loeb, 1993; 1994). One of the major contributors to the occurrence of monitoring errors is the manner in which patient data is currently relayed to the anesthesiologist. Physiological data displays are primarily visual in nature and follow a single-sensor-single-indicator paradigm (Görges & Staggers, 2007). The resulting sheer number of visual data displays puts the anesthesiologist at significant risk of data overload (Coiera, Tombs, & Clutton-Brock, 1996; Gaba, 1994; Sanderson, 2006), and this problem will likely only get worse as technological advances bring the introduction of yet more monitoring technologies to the operating room (Bitterman, 2006). The largely nonintegrated nature of these displays adds to the problem and requires considerable visual scanning, a behavior which anesthesiologists often cannot engage in due to competing demands for visual attention (Loeb, 1994). A common approach during periods of high visual demand is to periodically locate and sample parameters that anesthesiologists deem important or relevant, ignoring less important/relevant parameters. However, because anesthesiologists’ sampling patterns are often driven by incomplete or “buggy” mental models of the patient’s physiology (Cook & Woods, 1994), they may make flawed decisions about when and how often to sample the displays of various parameters. If a subset of physiological data displays are sampled too infrequently or too frequently (at the expense of other sources of information), there is an increased likelihood that critical information will be missed (Gaba, 1994).

One way to address these issues with current physiological data displays is to reduce the burden on the anesthesiologist’s visual and cognitive resources by employing other sensory modalities, such as audition or touch, either to convey information directly or to help capture and direct attention to the visual displays (Sanderson, 2006; Wickens, 2002; 2008). The employment of auditory signals, which have omnidirectional (i.e., can be perceived from any location and posture) and obligatory (i.e., no proximal means for completely eliminating the signal; we can't close our ears like we can our eyes) properties, is very effective for this purpose (Sanderson, 2006; Sarter, 2002; 2006). In a perfectly reliable system, a well-designed auditory alarm allows operators the freedom to visually scan data displays with much less frequency, thus allowing visual resources to be more fully committed to other tasks. Reliable auditory alarm systems can also serve to reduce some of the demand on cognitive functions such as memory and the need to rely primarily on endogenous mechanisms (e.g., driven by mental models and the goals and expectations of the operator) (e.g., Posner, 1980) to orient attention. They do this by providing more support for exogenous attention orientation (i.e., environmentally-driven attention capture) than do visual displays alone. These exogenous mechanisms are especially beneficial in supporting the detection of unexpected events (Remington, Johnson, & Yantis, 1992). The introduction of auditory alarms to the operating room has without doubt improved anesthetic monitoring performance. However there are a number of problems with these alarms that reduce their effectiveness, many of which collectively fit the description of what Woods classically refers to as *the alarm problem* (Woods, 1995; Woods, O'Brien, & Hanes, 1987).

The "alarm problem" in patient monitoring

The alarm systems associated with physiological data displays (hereon referred to as "physiological alarms") are plagued by high rates of false alarms (i.e., alarm signals activated when no true health event is present, due to electromechanical noise in sensor technologies which drive the alarm) and nuisance alarms (i.e., alarms that correctly activate when sensors detect a deviation in a system parameter, but the deviation is not related to a true health event and thus an alarm is not warranted) (Chambrin et al., 1999; Imhoff and Kuhls, 2006; Kesting, Miller, & Lockhart, 1988; Seagull and Sanderson, 2001). The prevalence of these alarms has the unintended effect of unnecessarily increasing the attentional load on the anesthesiologist (Cook & Woods, 1996), and distracting other OR personnel as well. Furthermore, most physiological alarms and notifications are rather uninformative, i.e., they simply announce that something has gone wrong but do not convey adequate information about the nature or urgency of the event, and often don't even specify the parameter they refer to (Dain, 2003; Imhoff and Kuhls, 2006; Meredith and Edworthy, 1995; Sanderson, 2006). Finally, since it is not uncommon to have multiple physiological systems affected by a single health event, a "cascading" sequence of multiple alarms - all competing for the anesthesiologist's attention - can make it difficult to find the root cause of the event. Combine these problems with the fact that the periods of densest alarm displays are also likely to be the time periods of highest cognitive load and task criticality for practitioners (when multiple systems fail concurrently), and alarms are just as likely to distract and disrupt problem-solving activities as they are to serve as an aid during these most critical periods (Cook & Woods, 1996; Woods, 1995). Given the list of problems with auditory alarms in the OR, it is not surprising that physiological alarms for many parameters are frequently silenced by default, or the threshold limits for activating the alarms are set wide enough to avoid sounding for all but the most extreme deviations of the parameter (Beatty & Beatty, 2004; Seagull & Sanderson, 2001; Watson, Sanderson, & Russell, 2004; Xiao, Mackenzie, Seagull, & Jaberi, 2000). Indeed, some have argued that the simple act of turning off these problematic alarms can lead to *improved* monitoring performance (Meyer & Bitan, 2002; Sorkin & Woods, 1985).

There are three common characteristics of physiological alarms that are likely responsible for, or exacerbate, many of the symptoms of the "alarm problem". The alarm systems are 1) threshold-based, 2) context-insensitive, and 3) binary in nature. Because the alarm algorithms are threshold-based, i.e., they do not activate until the level of a parameter surpasses a designated high or low threshold value, they cannot easily distinguish true health events from electromechanical noise in the equipment or artifacts such as those that may result from moving the patient. Consequently, the false alarm rate for these systems is notoriously high (as high as 90% (Imhoff & Kuhls, 2006)), and the resulting "crying wolf" effect (Lawless, 1994; Sorkin, 1988) may leave anesthetic personnel less prepared (and less likely) to act when a legitimate alarm state is realized.

To fully serve their intended purpose, threshold-based alarm systems require anesthetic personnel to decide on the most useful and appropriate threshold values for each physiological parameter and program them into the system. This is not a trivial task: high and low threshold values need to be set for upwards of 20 parameters which may be relevant for a given case. "Safe" ranges of values for some health parameters such as blood pressure and heart rate can vary considerably between patients, thus default alarm threshold values may not be appropriate. Further, the "safe" range for some variables can change for a single patient within a case. For example, some types of vascular surgeries require blood pressures that are much higher or lower than would normally be desired at different points in the same procedure, in order to minimize the risk of injury and/or assure that core organ systems receive sufficient blood perfusion during surgery (e.g., Roizen & Fleisher, 2002).

Another problem with threshold-based alarms is that they don't typically capture higher-order signal patterns which may be clinically relevant. Dangerous dynamics such as precipitous drops or rapid fluctuations in arterial blood pressure can occur without the absolute level ever exceeding threshold limits. These patterns can be just as, if not more, damaging to the patient's health as sustained high or low pressures (Rady, Ryan, & Starr, 1998). Additionally, such higher-order patterns are often present in the early stages of a health event and can serve as precursors to parameter levels exceeding an alarm threshold, which suggests that if anesthesiologists could be made aware of these patterns it may facilitate earlier recognition of and response to developing adverse events. For these reasons, "trend-based" alarm systems have been proposed and have shown to outperform threshold-based alarms in some areas (e.g., Fried & Imhoff, 2004; Haimowitz, Le, & Kohane, 1995; Schoenberg, Sands, & Safran, 1999), but ultimately show mixed results (Imhoff & Kuhls, 2006). Because they tend to have low sensitivity for detecting events by themselves, trend-based algorithms are best used as filters in connection with threshold-based alarms to help reduce the rate of false alarms by adding a layer of alarm intelligence.

The description of physiological alarms as being context-insensitive refers to the fact that the simple algorithms that are currently used to identify a potentially critical situation cannot take into account all of the contextual factors that the anesthesiologist may be aware of when judging the criticality of a situation. Physiological alarms are often classified by priority, in order to aid operating room personnel in deciding which events warrant shifting attention away from ongoing tasks, how urgently the shift is required, and which events to address first when there are multiple concurrent events. The priority of the alarm is then mapped to the urgency of the auditory signal. The problem is that there are frequently contexts in which the mapping between occurring events and the urgency communicated by the signal is inappropriate (Meredith & Edworthy, 1995; Woods, 1995; Xiao & Seagull, 1999). Events that urgently require attention in some contexts may be relatively benign in others. For example, a ventilator apnea alarm will sound when the system detects an absence of airflow over a period of time. The same alarm sounds if the patient has unexpectedly stopped breathing (which would be a big problem) or if the anesthesiologist has temporarily and intentionally disconnected the ventilator circuit from the patient (which is done frequently). The latter case does not constitute a false alarm; the alarm algorithm is performing as designed and is not negatively affected by system noise or artifact. Rather, this is an example of a nuisance alarm that attempts to direct attention to an event when the event does not warrant a shift in the operator's focus of attention (Woods, 1995).

One way to address issues with false and nuisance alarms and other problems related to the threshold-based and context-insensitive qualities of physiological alarms is to build algorithms that simultaneously analyze more than one monitored variable, thus allowing more and possibly converging evidence to be considered in determining the occurrence and severity of an event (e.g., Sukuvaara, Koski, Mäkivirta, & Kari, 1993). The alarms in clinical practice are predominantly univariate, but multivariate alarms have a growing presence in physiological monitoring systems, driven by a goal of improving alarm specificity through redundant sampling (Imhoff & Kuhls, 2006). For example, the rate of false alarms due to sensor error for heart rate irregularities can be reduced by logically combining data from multiple monitors, each of which can be used to identify a heartbeat: electrocardiogram, pulse oximetry, and arterial blood pressure monitors. Multivariate alarm algorithms might also be employed to allow events characterized by anomalies in multiple parameters to be more precisely identified. Taken a step further, multivariate systems could consider a higher level of contextual information, either input by anesthesiologists (e.g., "I'm going to disconnect the ventilator circuit now") or inferred from an integration of patient data. This ability could help reduce nuisance alarm rates (e.g., Görges, Markewitz, & Westenskow, 2009) and/or allow the alarm system to dynamically re-prioritize the alarms in order to avoid inappropriate urgency mapping of the alarm signal.

Finally, most physiological alarms are binary, i.e., they have only two states: on and off. Binary alarm algorithms are sufficient for capturing attention when a clinically relevant event occurs; however, they do little to support early problem identification and fast initiation of corrective actions. Because these events are often unexpected, anesthesiologists must first gather information about their nature (e.g., which parameter, is it high or low, HOW high or low, etc.) before they can respond, which takes time. It also requires the interruption of ongoing tasks, which is costly to overall task performance (Trafton & Monk, 2008). A binary alert leaves the anesthesiologist with an attentional catch-22: before they can decide whether the alarming event warrants a shift in attention from their ongoing tasks, they are required to shift their attention to learn about the event (Woods, 1995).

Another problem with binary, threshold-based auditory alarms is that after they serve their purpose to attract attention to a critical event, they continue to sound. Anesthesiologists must often take steps to manually silence the alarm(s) in order to proceed effectively with corrective actions (Watson & Sanderson, 2004; Xiao et al., 2000). “Silence all alarms” buttons which are included in most newer integrated anesthesia systems usually time out after a short amount of time, so the act of silencing the alarms may be required multiple times, especially if the problem-solving actions are particularly complicated or if the patient’s physiological response to the actions is sluggish. When operating room personnel are aware of the problem signaled by the alarm and are actively taking steps to resolve it, each repetition of the alarm is unnecessary and can be very disruptive and stressful. In this sense, these types of alarms are not unlike the aforementioned nuisance alarms.

One way to combat some of the problems with binary, threshold-based alarm systems may be the employment of multi-stage or graded alarm displays (Lee, Hoffman, & Hayes, 2004; Sorkin, Kantowitz, & Kantowitz, 1988; Woods, 1995). These displays, which can present an alarm signal that is proportional to the degree of threat, have been shown to improve attention allocation among tasks (Sorkin et al., 1988) and help an operator more accurately calibrate the extent to which they trust the validity of alarm systems (Lee, Hoffman, & Hayes, 2004; Lee & See, 2004). An alarm that announces a potentially developing situation which is not immediately critical allows the anesthesiologist freedom to decide whether or not to address the potential problem immediately. He/she can decide to temporarily postpone switching attention to complete an ongoing task or to see if the situation spontaneously improves. If and when the event reaches a more critical level, the anesthesiologist will be better prepared to respond to it because they will have been made aware of the event as it developed, resulting in them being in a “readied” state and reducing the need to search for more information before acting.

From alarms to improvements in “continuously informing” displays

Trend-based and multivariate algorithms, and graded alarms, are examples of efforts to address the alarm problems in patient monitoring. Thanks to these efforts, along with those to improve the reliability of sensor technologies and the robustness of signal filters and statistical data processing algorithms (e.g., Imhoff & Kuhls, 2006), physiological alarms are becoming more reliable and useful as a diagnostic aid. Interestingly and paradoxically, these developments may not be entirely positive for the practice of physiological monitoring. One consequence of employing highly but imperfectly reliable automation (such as an alarm system) is that humans tend to place an excessive amount of trust in the system to detect problems and capture their attention when it is necessary to do so (Lee & See, 2004; Parasuraman & Riley, 1997). The more reliable the patient monitoring alarm system becomes, the more complete the reliance on the system will be to detect and announce any problems (McGuirl & Sarter, 2006; Mosier, 1997; Sarter & Schroeder, 2001). This means anesthesiologists will become less aware of the behavior of the physiological parameters which are handled by the alarm system. When something eventually does go wrong that is not handled effectively by the system (and with imperfect automation, something always does), the anesthesiologist will be left “out of the loop”. They may miss the problem entirely or be left unprepared to act quickly and effectively to address it (Bainbridge, 1983; Mosier, 1997).

As research efforts continue to improve the design of physiological alarms, it is not likely that these improvements will fix all the problems with the systems, at least not in the foreseeable future. Therefore, rather than focusing exclusively on alarm design it may be necessary to fundamentally reconsider the way in which patients’ physiological data is relayed to the anesthesiologist. The primary function of the alarm systems is to assure that anesthesiologists are made aware of patient health data (which the system deems relevant) when they may otherwise not be aware of it. This suggests that if an anesthesiologist’s real-time awareness of these data can be improved, the need to rely on alarms can be reduced. Such an improvement may be realized through the redesign of patient physiological data displays. These redesigns would need to accomplish some or all of the following 6 goals:

1) *Improve, or at least not reduce, detection of abnormal levels in a physiological parameter*
2) *Support earlier diagnosis of and response to developing events, ideally before parameters reach critical levels*
3) *Support more precise management of displayed parameters within prescribed ranges (through pharmacological or other physical interventions)*
4) *Reduce competition for visual resources*
5) *Reduce, or at least not increase, overall cognitive workload*

6) *Provide better support for attention and task/interruption management by assisting anesthesiologists in making informed decisions about whether, and when, to switch attention between physiological monitoring tasks and other ongoing tasks*

There are three promising approaches to display redesign which are prevalent in the literature and satisfy at least some of the goals listed above: advanced graphical displays, head-mounted displays, and nonvisual "continuously informing" displays.

Advanced graphical displays. Advanced graphical displays represent a departure from the strictly "single-sensor-single-indicator" paradigm of current physiological data displays and usually involve an integrated graphical representation of multiple patient health parameters. Examples include configural, or object, displays which use separate dimensions of a visual display object to relay different, but often related streams of data (e.g., Bennett & Flach, 1992; Drews & Westenskow, 2006). The dimensions can be configured in such a way that anomalies in any individual dimension, or patterns across multiple dimensions, are perceptible as emergent features of the overall display (e.g., Pomerantz & Pristach, 1989). Configural and emergent features displays representing pulmonary (Wachter, Johnson, Albert, Syroid, Drews, & Westenskow, 2006), cardiovascular (Albert, Agutter, Syroid, Johnson, Loeb, & Westenskow, 2007), hemodynamic (Blike, Surgenor, Whalen, & Jensen, 2000), drug level (Charabati, Bracco, Mathieu, & Hemmerling, 2009; Drews, Syroid, Agutter, Strayer, & Westenskow, 2006), and combinations of these data (Michels, Gravenstein, & Westenskow, 1997; Zhang et al., 2002) have shown to improve patient monitoring performance in a number of simulation studies, when compared to more traditional display methods. Most of these studies show improved detection and diagnosis of adverse health events and earlier treatment response, when compared to traditional visual displays (Drews & Westenskow, 2006; Görges & Staggers, 2007). The need for visual scanning is reduced by co-locating related data, and cognitive workload is reduced by transforming the tasks of comparing and combining data from multiple parameters to the task of recognizing simple features or patterns such as object size, shape regularity, and symmetry (Drews & Westenskow, 2006). It is important to note, however, that the majority of studies evaluating the various advanced graphical displays are conducted without imposing any significant workload by secondary tasks (Görges & Staggers, 2007; Sanderson, Watson, & Russell, 2005); therefore, performance effects of these displays that are related to task and interruption management have yet to be demonstrated.

Head-mounted displays. A fundamental limitation of advanced graphical displays is that, while they may reduce scanning and information access costs related to scanning multiple displays, they still require sampling a visual display, which may be difficult, especially at busy times (Loeb, 1994; Sanderson et al., 2005). Head-mounted display (HMD) technology can help by keeping patient data displays always within the field of view of an anesthesiologist, thus removing the need for postural or positional changes to view relevant data. Some evaluation studies have found faster detection of critical patient health events while wearing a HMD, when compared to event detection with traditional visual displays (e.g., Via et al., 2003). However, this is not a universal finding (e.g., Sanderson et al., 2008), and an advantage might only be consistently observed in cases where the anesthesiologist is physically constrained and unable to immediately glance toward traditional visual displays (Liu et al., 2009). When equipped with a HMD, anesthesiologists spend more time looking towards the patient and overtly switch visual attention to other (traditional visual) displays less frequently (Liu et al., 2009; Ross, Ormerod, Hyde, & Fine, 2002). This suggests that HMDs may better support task and interruption management, but again this has yet to be clearly demonstrated empirically. Finally, there are justified concerns that anesthesiologists wearing HMDs may be susceptible to missing critical events even though they are within the field of view due to differences in the plane of focus (Liu et al., 2009) or to attentional phenomena such as "inattentional blindness" (e.g., Mack & Rock, 1998). The effects of inattentional blindness have been clearly demonstrated in other domains that impose comparable visual and cognitive demands on an operator. In aviation, for example, pilots using head-up displays (HUDs) have missed runway incursions by other aircraft as they were landing, even when they were clearly visible and within the same field of view as the display (e.g., Haines, 1991; Wickens & Long, 1995).

Nonvisual displays. Nonvisual "continuously informing" displays – those that present a continuous or semi-continuous stream of data via auditory or haptic signals – have the potential to best satisfy the 6 stated goals for improved physiological display technologies. These displays have the advantage of reducing competition for visual resources. If designed properly, the auditory or tactile information can be processed in parallel with ongoing visual tasks, thus reducing the effective overall attentional load and improving multitask performance (Wickens, 2002; 2008). They share many of the benefits of graded/multistage alarms, allowing anesthesiologists to detect and track developing events earlier, which leads to more accurate diagnoses of problems and more timely responses. Continuously informing displays also allow the anesthesiologist to decide whether and when it is most appropriate to switch attention from an ongoing task to address deviations in health parameters, thus improving interruption and task management.

To date, researchers have focused solely on one nonvisual modality – audition – for the continuous display of patient data. “Sonification” displays are auditory displays that map numerical values or relations in patient data to values or relations in one or more auditory signal dimensions (Kramer, 1994). The gold standard for clinical sonification displays, in fact, the display that is consistently recognized as the most important and useful of all patient data displays for detecting and diagnosing critical health events, is the variable-tone pulse oximetry display (James, 2003; Webb et al., 1993). This display consists of a periodic tone that sounds with each heartbeat/pulse, thus mapping tone rate to heart rate. The tone pitch is mapped to blood-oxygen saturation, thus a descending pitch between heartbeats clearly and intuitively communicates falling blood oxygenation. Currently, this remains the only sonification display to have taken hold in clinical practice; however, researchers inspired by the success of this display have proposed and evaluated sonifications of other patient data, including blood pressure and various respiration measures (Fitch & Kramer, 1994; Loeb & Fitch, 2002; Sanderson et al., 2008; Seagull, Wickens, & Loeb, 2001; Watson & Sanderson, 2004). Evaluation studies have shown mixed results. Some sonifications outperformed traditional visual displays in the detection rate, identification, and/or response time to physiological events (Fitch & Kramer, 1994; Sanderson et al., 2008). Others found that sonification alone led to worse monitoring performance than visual displays alone (Seagull et al., 2001; Loeb & Fitch, 2002), but that a combined visual + sonification display led to the best overall performance (Loeb & Fitch, 2002). Still others found no differences in monitoring performance between sonification and traditional visual displays (Watson & Sanderson, 2004). Importantly, some studies required a demanding visual task to be conducted concurrently while monitoring patient health, imposing a load on visual resources that is representative of an anesthesiologist’s task set in the clinical setting. Performance on this secondary task improved in the sonification-only display conditions (Seagull et al., 2001; Watson & Sanderson, 2004), demonstrating the potential to support multitasking by engaging separate display modalities.

The success of the pulse oximetry sonification may be due to its ability to communicate the level of an important variable in a way that minimally competes for attention (e.g., Seagull et al., 2001; Watson & Sanderson, 2004). Because the auditory stream is nearly continuous, the sonification signal lacks the attention-capturing qualities that a discrete-onset auditory alarm would. When the signal is relatively unchanging, it exists "in the background", providing the anesthesiologist with peripheral awareness of blood-oxygen saturation and heart rate. In this mode of operation, the pulse oximetry sonification can be described as an "ambient" display (Ishii et al., 1998; MacLean, 2009, Wisneski et al., 1998). Changes in the pulse oximetry signal, especially when large and unexpected, can cause a natural transition of the signal from peripheral to focal awareness. Because the dimensions of the auditory signal are so well mapped to the represented variables, the anesthesiologist immediately knows the nature of the physiological dynamic that triggered the signal change, and can estimate the magnitude or severity of the dynamic to decide if it warrants a shift in attention to perform a corrective action.

These qualities of the pulse oximetry display, which other sonification designs strive to emulate, reflect those described by Woods (1995) as characteristics of a display that supports "preattentive reference". Coined after the concept of preattentive processing described in attention research (e.g., Broadbent, 1977), the term preattentive reference describes displays that (Woods, 1995):

1) *are capable of being processed in parallel with an operator's ongoing tasks without interference;*
2) *include partial information on what attention-directing signals refer to, so the operator can infer whether the signal warrants a shift in attention or not; and*
3) *can be assessed in a mentally economical way (without requiring an act of focal attention).*

One could argue that when compared to the pulse oximetry display, the relative lack of success for other sonification physiological displays is due to characteristics which make them less effective in regard to one or more of the preattentive reference qualities listed above. The problem may be related to the sonification configuration, i.e., the number of auditory streams used to carry information (e.g., Anderson & Sanderson, 2004; Kramer, 1994). Preattentive reference is best supported when a signal can be assessed in a mentally economical way; this suggests keeping a sonification display relatively simple, using one or few auditory streams to relate the data, and relying as much as possible on natural mappings between signal modulations and the underlying data to facilitate interpretation with minimal cognitive engagement (e.g., Sanderson, Liu, & Jenkins, 2009). The pulse oximetry display includes data related to just two parameters (heart rate and blood oxygenation) and employs two very naturally-mapping signal modulations (tone rate and pitch) to relate this data. In contrast, the studies mentioned previously sonified considerably more patient variables: five (Sanderson et al., 2008; Watson & Sanderson, 2004), six (Fitch & Kramer, 1994; Loeb & Fitch, 2002; Seagull et al., 2001), and one design supports as many as eight (Fitch & Kramer, 1994). Including all of these variables into a single auditory stream would create a display that is very complex, and it would be very difficult to assign each variable to a signal dimension that exploits its most natural mapping. It is not surprising therefore that the designers of each sonification design discussed here separated the sonified variables into two auditory streams, which allowed them to group related variables in each and exploit natural mappings as much as possible, thus minimizing overall display complexity.

While separating the sonified data into multiple auditory streams may help the ability to assess signals in a mentally economical manner, it poses a greater challenge to another characteristic of preattentive reference: the ability to process data in parallel with ongoing tasks. Perceptual and cognitive interference will be greater when numerous data streams engaging the same sensory modality and sharing a similar structure are processed concurrently (Wickens, 2002; 2008). Humans in a divided-attention state (the state benefiting the most from preattentive reference displays) clearly show a reduced ability to detect an irregularity in one of multiple complex auditory signals as the number of streams increases (Anderson & Sanderson, 2004; Brochard, Drake, Botte, & McAdams, 1999). Thus, in the design of sonification displays, care must be exercised to avoid overloading the auditory channel or the same monitoring problems that plague existing visual displays will be recreated.

The potential for tactile displays to support patient monitoring

The above challenges associated with sonification design are part of a considerably larger problem in clinical settings: the operating room is a very noisy place. Auditory clutter created by unnecessary auditory alarms (e.g., Dain, 2003; Hodge & Thompson, 1990) is one part of the problem. Equipment such as surgical drills/saws and vacuum pumps, vocal communications between OR personnel and from outside sources via intercom, and even music playing from an OR stereo all combine to create a noise level that has been compared to standing alongside a busy highway (Hodge & Thompson, 1990; Ulrich et al., 2008). This high level of background noise may lead to auditory masking of alarm signals, causing them to be missed or misinterpreted (Lacherez, Seah, & Sanderson, 2007; Momtahan, Hetu, & Tansley, 1993; Sanderson et al., 2009), especially by anesthesiologists with moderate age-related hearing loss (Wallace, Ashman, & Matjasko, 1994). Confirming this concern, 8 of 34 anesthesiologists who responded to an online survey that was conducted as part of the present research listed "background noise in the OR" in a free-response question about factors that significantly affect their monitoring performance (see Chapter 3).

Another issue with auditory displays in the OR is their ubiquitous and socially inclusive nature (e.g., Sanderson et al., 2009). Free-field auditory displays are appropriate for some parameters, such as pulse oximetry, which are important and frequently relevant to the tasks of not only the anesthesiologist but other OR personnel as well (e.g., Sanderson, 2006). However, there are many other parameters – such as blood pressure and respiratory measures – that are much more relevant to anesthesiologists than other personnel under normal conditions. To avoid irritating and unnecessarily distracting these personnel with auditory displays (a motivating factor in the silencing of alarms), it would be beneficial to have a non-visual display option that can communicate information in a more private manner.

Tactile displays – which communicate via the sense of touch – represent a new and promising means to achieve this goal. Some forms of tactile display, especially the presentation of information through coded patterns of vibrations on the skin (i.e., *vibro*tactile diplays), have seen growing interest in recent years (e.g., Brewster, Wall, Brown, & Hoggan, 2008; Gallace, Tan, & Spence, 2007; Jones & Sarter, 2008; MacLean, 2008b; Pasquero, 2006). The two main reasons are 1) the need to support communication with human operators in complex work environments where the visual and auditory channels have reached saturation and 2) the improved commercial availability of sophisticated tactile display technologies (e.g., Jones & Sarter, 2008; Pasquero, 2006).

The sense of touch has a set of affordances which uniquely qualify it for displaying some types of information in a clinical setting (e.g., Sanderson, 2006). Like audition, touch is obligatory and omnidirectional, such that any person equipped with a tactile display can receive its signals regardless of the spatial orientation of attention. Unlike audition, touch more easily supports privatization of information because it is a proximal sense, i.e., only those who are in physical contact with tactile display devices have the ability to receive its message. This means system-initiated tactile messages can be reliably communicated to the anesthesiologist in a way that does not annoy or unnecessarily distract other OR personnel.

Touch also features a superior ability to process stimuli that require discrimination in both the spatial and temporal signal dimensions. The spatial acuity of tactile perception is at least as high as audition (depending on body location) and the abilities to judge the temporal durations, rates, and rhythmic patterns of stimuli are comparable for the tactile and visual channels, and favor touch in many cases (Geldard, 1960; Sherrick & Cholewiak, 1986; Welch & Warren, 1986). This makes it an ideal channel for relating information that naturally maps to directions, such as “up” or “down” for increasing/decreasing levels of a parameter, and also for information that has a temporal quality, like heart rate or respiration rate. An important signal dynamic which can be well-represented by a tactile display is the rate of change in a parameter, which combines both spatial and temporal properties. For example, a tactile display may efficiently relate how quickly blood pressure is rising or falling, which is a significant factor in deciding the clinical relevance of the pressure change.

Recognizing this potential, one group of researchers at the University of British Columbia has recently begun designing and evaluating tactile displays for the purposes of supporting physiological monitoring (Barralon, Dumont, Scharz, Magruder, & Ansermino, 2009; Barralon, Ng, Dumont, Schwarz, & Ansermino, 2007; Ford et al., 2008; Ng, Barralon, Schwarz, Dumont, & Ansermino, 2008; Ng, Man, Fels, Dumont, & Ansermino, 2005). Primarily driven by a goal of reducing OR noise levels and, as an extension, reducing the likelihood of missing alarms, this research group demonstrated some encouraging results in the tactile display of information related to heart rate (Ng et al., 2005), respiratory measures (Ford et al., 2008), and critical events related to four (Ng et al., 2008) or six (Barralon et al., 2009) different physiological measures through multiparameter displays. In each case, these displays were designed to extend the informativeness of existing auditory alarms by encoding additional information into the tactile signal, such as the severity of an alert and the direction of change in a parameter. Generally, study findings have shown that the tactile messages can be learned fairly quickly. The identification of and response to tactile alarms is comparable and, in some cases, better than performance with standard auditory alarms. And participants rated the tactile displays as useful and, in some cases, preferable over auditory displays.

The work of this pioneering group still leaves a number of questions and directions for tactile display design and evaluation in the context of physiological monitoring. First, the vibration displays designed by the UBC group present informative and graded alarm signals, but they are still threshold-based alarm systems and thus subject to the same set of alarm problems as their auditory counterparts. Tactile versions of the "continuous informing" physiological displays remain relatively unexplored. An early pilot study by this group (Ng & Man, 2004) investigated one very simple form of continuous display – presenting a 200-ms vibration pulse with each heartbeat to convey heart rate information – but poor performance results and high annoyance ratings (likely due to the frequency of presentation: as often as twice per second) led them to discontinue this pursuit in favor of the alarm displays.

Second, for the more complex multiparameter displays, critical events are quickly and reliably detected but the interpretation of, and response to, the signals could be improved. Alarm misidentifications were somewhat frequent (up to 20%), and times spent decoding the signals before initiating a response were long (roughly 5 to 10 s for different conditions) (Barralon et al., 2007; 2009; Ng et al., 2008). These findings may be partially attributable to mapping issues, which could be a consequence of the number of parameters and amount of information encoded in the display. Confusions about which physiological parameter was involved in a critical event may be due to their somewhat arbitrary associations with tactile device locations on a belt around the waist/abdomen (Barralon et al., 2007; 2009; Ng et al., 2008). The belt-mounted display also employed rather nonintuitive encoding methods for communicating trend direction, mapping "increasing" and "decreasing" to either the duration of signal pulses (Barralon et al., 2009; Ng et al., 2008) or the "roughness" of signal waveform (Barralon et al., 2007). Another display prototype involved an array of vibrating devices on the back, and the physiological parameter associated with a critical event was communicated by a vibration sequence which effectively traced the first letter of that measure on the array (Barralon et al., 2009). This method showed some promising results including higher identification accuracies, but the time required to present the full pattern including tracing the letter took several seconds (mean of 4.4 s), and participants took considerable time processing the signal after it was fully presented (mean of 5.6 s) before initiating a response. Supporting more natural information-to-signal dimension mappings by using, for example, metaphorically-derived patterns (e.g., MacLean, 2008a; 2008b) may improve the accuracy and speed of their interpretation.

Finally, with the exception of the belt-mounted respiratory alarm display (Ford et al., 2008), these displays were evaluated in contexts that did not include realistic physiological monitoring scenarios nor multitask sets representative of those of an anesthesiologist. Further, evaluations were often conducted with non-practitioners (i.e., study participants were not medically trained). These limitations are highlighted by the researchers and can be explained by their stated goals, which focused primarily on the detection and interpretation of coded tactile messages, and not on physiological management or multitask performance. Nonetheless, they remain a concern in drawing conclusions about the usefulness of these displays in a clinical setting. Note that this concern is one that has been voiced for many studies designed to evaluate advanced physiological display technologies (Görges & Staggers, 2007; Sanderson et al., 2005). The true clinical and operational benefit of these displays can only be realized when they are tested with anesthesiologists under an appropriate task load, and evaluated according to measures that are directly related to patient outcomes.

The case for a "tactification" physiological data display system

This dissertation describes research efforts to develop and evaluate a "continuously informing" vibrotactile display with the goal of supporting attention and task management of anesthesiologists in a clinical setting, as well as operators in other data-rich event-driven domains. This work is novel in that it investigates the potential for a new direction in tactile displays: the real time, (semi)continuous presentation of (patient) data via coded vibration patterns. Such a "tactification" display complements earlier work on auditory sonification displays of physiological data, and has the potential to lead to superior performance for anesthesiologists (and other operators) when their visual and auditory modalities are heavily loaded due to concurrent task demands. Chapters 2, 3, 4, and 5 describe research activities that needed to be conducted first to answer theoretical and practical questions about supporting attention and task management with tactile displays. The findings from those studies served as input to the final tactification display designs. Chapter 5 describes these display designs and an evaluation study conducted with anesthesiologists in a high-fidelity patient simulator.

The central theoretical contribution of this line of research is that it provides greater insight into the properties, affordances, and limitations of the tactile modality and communications via tactile displays. In particular, this work describes how tactile displays can be used to relay data in a way that can keep an operator continuously informed of the state of multiple system parameters and thus support them in allocating attentional resources and balancing workload demands. A theory-based approach was taken in designing the tactile displays to achieve at least some of the goals of preattentive reference displays, paraphrased below from Woods (1995):

1) *Display signals can be reliably perceived and interpreted while minimally interfering with processing required for concurrent tasks.*

   When attempting to process data streams related to separate tasks in parallel, interference can occur at various processing stages (e.g., perception, cognition, and response stages; Wickens, 2002; 2008). As is the case with processing via any sensory modality, the perceptual and cognitive stages are the most relevant when discussing the introduction of tactile information processing to a concurrent task set.

   One of the primary motivations for employing tactile displays in environments that heavily load the visual and/or auditory channels is to minimize processing interference at the perceptual stage (Wickens, 2002; 2008). However, a complex tactile display – one that includes multiple streams of data, e.g., related to multiple health parameters – may suffer from intramodal perceptual and attentional interferences. One such interference may be described as a tactile analog of visual change blindness, i.e., the surprisingly poor ability humans have detecting even large changes in a visual scene when the changes occur at the same time as a brief masking visual stimulus, such as a flash of light or blank screen (e.g., Rensink, O'Regan, & Clark, 1997). Chapter 2: Part A of this dissertation describes a study that investigated how some vibrotactile signals and patterns may induce a version of this effect in regard to the detection of changes in one very important dimension in tactile display design: vibratory stimulus intensity.

   "Downstream" of perception, task-related information is interpreted at the cognitive processing stage. Processing interferences can be observed at this stage to the extent that concurrent tasks require similar processing codes (Wickens, 2002; 2008). For example, reading text and listening to speech engage separate sensory modalities, but require the same verbal processing code resources, hence are very difficult to effectively conduct in parallel. When this research effort was started, it was not yet known to what extent the dimension of processing code may affect the ability to process a complex tactile display message in parallel with ongoing visual/auditory task information. To address this, a pair of studies was conducted which investigated how tactile displays that use spatial location (i.e., different locations on the body) to encode information, and displays that use nonspatial encoding methods (e.g., temporal patterns), suffer from interferences at the cognitive processing stage when visual tasks require the same resources (see Chapter 2: Part B).

2) *Signals include partial information on what attention-directing signals refer to, so the operator can infer whether the signal warrants a shift in attention or not.*

A display that effectively supports this goal must a) encode partial information about a secondary task in a way that can be reliably detected and accurately interpreted, and b) communicate information that supports efficient task switching, i.e., switching attention to the secondary task when it is necessary to do so, but minimally affecting performance on an ongoing primary task when a switch is not necessary. Taken together, displays that communicate partial information about a secondary task best support this goal when the highest combined performance levels can be seen in primary *and* secondary tasks.

For a secondary task that requires monitoring and managing a dynamic variable (such as an anesthesia patient's blood pressure), the two most relevant types of information in the decision of whether to switch attention from a primary task to manage the variable are: a) the level/state of the variable; and b) the direction and rate of change of variable levels over a recent time interval. These types of information are reflected in the two types of physiological monitoring alarm systems: threshold-based and trend-based alarm algorithms, respectively.

Based on responses to some questions of an online survey by anesthesiologists (see Chapter 3), three prototype vibrotactile displays were designed to relate information regarding 1) blood pressure level; 2) direction and rate of change in the level; and 3) a combined presentation of both the level and direction/rate of change of pressures. Also following input from the survey, an experimental simulator was developed which included a primary task modeled after some of the most challenging aspects of anesthesia induction, and a secondary blood pressure monitoring and management task. Chapter 4 describes a study set in this simulator which investigated how the amount and types of information encoded in the tactile displays affected the combined performance in both the primary (induction) and secondary (blood pressure management) tasks.

3) *Signals can be assessed in a mentally economical way.*

After identifying the types of information that best support task management in a multitask set (Chapters 4 and 5), it is then important to find ways to relate this information via display signals which can be interpreted with minimal cognitive effort. One strategy to achieve this is to rely as much as possible on natural mappings between the represented information and the signal modulations used to relate it (e.g., Norman, 2002). As discussed earlier, the tactile channel is uniquely suited for communicating information that has either/both spatial and temporal qualities, due to its relatively high discrimination ability in those dimensions (Geldard, 1960; Welch & Warren, 1986). Because physiological data can often be described in a way that intuitively maps to those same dimensions (e.g., high/low levels in a physiological parameter can map well to up/down spatial locations, and rates of change in these levels can map well to temporal presentation patterns), the final tactification designs described in Chapter 5 exploited these mappings. Responses to survey questions (Chapter 3) also provided input as to how to naturally relate some signal levels and dynamics.

Where it was feasible to do so, redundant signal modulations were used to relate data; this increases the chances of the most natural signal mappings representing the data. For example, the tactification displays of blood pressure redundantly related pressure levels by spatial location (higher on the arm for higher pressures) and vibration intensity (higher intensity for higher pressures). Redundant modulations such as these also help to minimize processing interferences at the cognitive stage (Chapter 2: Part B), e.g., by facilitating interpretation of the signal by engaging either the spatial or nonspatial processing codes, whichever interferes less with the processing required for ongoing tasks.

Finally, when multiple different systems, or multiple dynamics within the same system, are represented via tactile displays, efforts must be taken to help the observer distinguish them with minimal effort. One way to do this is to use symbolic or metaphorical mapping of the signal to facilitate its identification (MacLean, 2008a; 2008b; 2009). For example, the tactification designs in Chapter 5 relate blood pressure by vibrations on the upper arm (where a blood pressure cuff would normally be affixed), with a signal that "feels" like a heartbeat. This type of metaphorical representation helps the observer easily determine that changes in this signal are related to blood pressure rather than respiratory measures.

In addition to adding to the knowledge base in tactile information processing and informing the design of tactile displays, this work helps advance models of the structure of cognitive resources, such as Multiple Resource Theory (e.g., Wickens, 2002; 2008). This research investigates human performance under conditions for which these models are not yet well-defined (Wickens, 2008): those that involve processing of continuous tactile signals; and performance in task sets that involve the continuous processing of task-relevant data in three sensory modalities (vision, audition, and touch) concurrently.

This work also has important practical implications. The efforts described in this dissertation to design a tactile display to support physiological monitoring directly address a major problem in healthcare: the prevalence of preventable monitoring errors. These errors can be explained, to a large extent, by the current manner in which physiological data is displayed in the clinical setting. The tactification displays described in Chapter 5 represent a fundamental change in the manner in which patient data is related to the anesthesiologist, and a promising approach to reduce monitoring errors.

The tactification displays communicate real-time levels of three important physiological parameters: arterial blood pressure, respiratory tidal volumes (the volume of gas in each breath), and end-tidal carbon dioxide (the partial pressure of CO_2 in each exhaled breath). These data are relayed via semicontinuous patterns of vibrations: every three seconds for blood pressure; with every breath for the two respiratory measures. These displays represent a potential improvement over existing physiological display technologies by addressing each of the 6 following goals:

1) *Improve, or at least not reduce, detection of abnormal/critical levels and dynamics in a physiological parameter*

Because the tactile sense is omnidirectional, the detection of a problem state which is communicated by a tactile display is not limited by the orientation of sensory organs, as is the case with visual displays. Tactile displays can be at least as effective as auditory displays at announcing an event in an alarm format (i.e., only vibrating/sounding when an abnormal level is realized). But in a more continuous format, a tactification display may be more attractive than a sonification display by virtue of being a private presentation (i.e., the signal can be presented to only the anesthesiologist/other relevant personnel). Because the tactile signal will not annoy/distract other personnel when it is unnecessary for them to be exposed to it, the tactification display may be more likely to be used, and thus abnormal parameter levels more likely to be detected.

Comparing (auditory, visual, or tactile) alarms to a tactification display should also show similar performance in supporting detection of abnormal parameter levels. By exploiting natural mappings between each parameter level and modulations of the vibration signals, abnormal levels may be identified with minimal cognitive effort. Further, the tactification display may be superior to alarm displays in this sense, by supporting identification of problem states related to parameter dynamics such as erratic (but not abnormally high or low) blood pressures or irregular respiration rates.

2) *Support earlier diagnosis of and response to developing events, ideally before parameters reach critical levels*

Most current physiological alarm systems are designed to alert the anesthesiologist when a parameter crosses a high or low threshold value; however, at this point the patient may already be in danger of suffering lasting health effects. If the event is unexpected and the anesthesiologist is not prepared to act immediately, the patient could suffer further while the anesthesiologist diagnoses the problem and decides on corrective actions. Therefore, earlier detection of trends which could lead to these critical states is beneficial: the anesthesiologist may be able to take corrective actions before the critical state is realized, or can at least be better informed and prepared to act when the parameter levels do reach dangerous levels. By communicating realtime parameter levels semicontinuously, the tactification display will improve detection of these earlier trends, as anesthesiologists can notice changes in the level. Further, the speed with which these changes occur can help the anesthesiologist predict how quickly a problematic situation may arise and allow them to plan accordingly around ongoing task activities.

3) *Support more precise management of displayed parameters within prescribed ranges (through pharmacological or other physical interventions)*

Tactification displays can support the earlier detection of and response to events which can ultimately result in parameter levels outside of prescribed ranges. Therefore, by addressing these dynamics sooner, both the distance outside of these ranges and the amount of time spent outside of the ranges for parameter levels will be less.

Because pharmacological interventions (i.e., administering drugs to "correct" a parameter) can take some time to realize their full effects, a tactification display also provides feedback while corrective efforts are underway. This is useful, for example, to help recognize cases of under- or overdosing. A drug overdose can cause the parameter being corrected to overshoot the desired range; earlier recognition of these cases can also result in more precise management of the parameter.

4) *Reduce competition for visual resources*

Anesthesiologists are responsible for multiple concurrent tasks during the anesthetization of a patient which place high demand on visual attention. They cannot afford to visually monitor physiological parameters at all times to ensure they are always within desired ranges, or performance on the concurrent tasks will suffer. By communicating some parameters via the tactile channel, the need to visually sample displays for those parameters is reduced. The result is that more time can be spent visually focusing on the patient and on other care-related tasks, such as programming drug infusion machines.

Sonification displays can reduce competition for visual resources, but likely do so at the cost of increased competition for auditory resources. In contrast, tactification displays might be able to keep the anesthesiologist more continuously informed without negatively affecting the ability to process the pulse oximetry display and auditory alarms. By keeping auditory clutter in the operating room (and the associated unnecessary distraction and annoyance) to a minimum, the entire team of OR personnel may benefit from this display configuration. Auditory signals can be reserved for events which are only of the highest importance and/or relevant to the entire OR team, while tactile signals can be used to privately communicate information that is usually only relevant to the anesthesiologist.

5) *Reduce, or at least not increase, overall cognitive workload*

The tactification displays can support at least some of the qualities of preattentive reference. These qualities allow the state of parameters to be consciously sampled with minimal cognitive effort. Additionally, they help the anesthesiologist recognize abnormal states effectively and through perceptual mechanisms (such as an abrupt change in the position or intensity of the vibration signal) without engaging focused attention. As a result, under normal conditions the signal can be "tuned out", allowed to exist in the background of consciousness, since significant changes in the signal (those which likely signify a critical event developing) will be reliably detected.

6) *Provide better support for attention and task/interruption management by assisting anesthesiologists in making informed decisions about whether, and when, to switch attention between physiological monitoring tasks and other ongoing tasks*

With a tactification display, anesthesiologists can endogenously orient attention to sample the state of the displayed parameters whenever they choose; they are less limited by physical constraints such as postures that make visual sampling of displays difficult or impossible. Additionally, the tactification display provides a more efficient exogenous attention orientation mechanism than existing alarm systems. By providing a signal that naturally maps the severity of a developing event to its salience, it can remain below the threshold for attention capture until a dynamic is significant/severe enough to attract attention. The anesthesiologist can then make an initial assessment of the event to decide whether and when it is most appropriate to switch tasks to attend to the problem. Thus, fewer unnecessary interruptions will occur because the anesthesiologist can make a more informed decision about the need to switch the focus of attention, taking into account contextual information that automated alarm systems currently cannot when judging an event's clinical relevance.

In addition to supporting a reduction in the number of interruptions, the time taken to tend to an interrupting task (e.g., taking actions to manage a physiological parameter) can be shorter with tactification displays as well. By being better informed about a developing health event, anesthesiologists can be better prepared to act quickly when it is necessary to switch tasks to address it. The event is then addressed more quickly, and an anesthesiologist can return to the interrupted task sooner and with less performance cost associated with task-switching (Trafton & Monk, 2008).

In summary, the research activities described in this dissertation constitute a significant contribution to theories of tactile and multimodal information processing and the structure of cognitive resources. A theory-based approach was taken to the design of a new type of display, one that can keep an operator more continuously informed about the state of a monitored system through patterns of vibrations applied to the skin. This type of display has unique properties that approach the qualities of preattentive reference displays, and can effectively support a human operator in the management of attention while conducting multiple concurrent tasks. By designing these displays specifically to support the attention of anesthesiologists while they perform a task set that is representative of what they encounter in the clinical setting, the results of this work can inform display design to improve physiological monitoring performance. In addition to benefitting anesthesiologists, this work may ultimately also benefit the healthcare business and community of healthcare providers by suggesting ways to reduce a major form of preventable medical error, and reducing the costs associated with rectifying these errors. Finally, the consumers of the healthcare industry – society at large – can benefit when healthcare providers such as anesthesiologists are able to care for their patients more effectively, thus resulting in less occurrence, and more *prevention* of preventable errors.

REFERENCES

Albert, R.W., Agutter, J.A., Syroid, N.D., Johnson, K.B., Loeb, R.G., and Westenskow, D.R. (2007). A simulation-based evaluation of a graphic cardiovascular display. *Anesthesia and Analgesia, 105*(5), 1303-1311.Bainbridge, L. (1983). Ironies of automation. *Automatica, 19*(6), 775-779.

American Society of Anesthesiologists (ASA). (1986; reaffirmed 2005). *Standards for Basic Anesthetic Monitoring*. Available at http://www.asahq.org/publicationsAndServices/standards/02.pdf

Anderson, J. & Sanderson, P.M. (2004). Designing sonification for effective attentional control in complex work domains. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*, 840-844.

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19*, 775-779.

Barralon, P., Dumont, G., Schwarz, S.K.W., Magruder, W., & Ansermino, J.M. (2009). Comparison between a dorsal and a belt tactile display prototype for decoding physiological events in the operating room. *Journal of Clinical Monitoring and Computing, 23*, 137-147.

Barralon, P., Ng, G., Dumont, G., Schwarz, S.K.W., & Ansermino, M. (2007). Development and evaluation of multidimensional tactons for a wearable tactile display. *Mobile HCI'07*, 186-189.

Beatty, P.C.W. & Beatty, S.F. (2004). Anaesthetists' intentions to violate safety guidelines. *Anaesthesia, 59*(6), 528-540.

Bennett, K.B. and Flach, J.M. (1992). Graphical displays: Implications for divided attention, focused attention, and problem solving. *Human Factors, 34*(5), 513-533.

Bitterman, N. (2006). Technologies and solutions for data display in the operating room. *J Clinical Monitoring and Computing, 20*, 165-173.

Blike, G.T., Surgenor, S.D., Whalen, K., and Jensen, J. (2000). Specific elements of a new hemodynamics display improves the performance of anesthesiologists. *J of Clinical Monitoring and Computing, 16*(7), 485-491.

Brewster, S., Wall, S., Brown, L., & Hoggan, E. (2008). Tactile displays. *The Engineering Handbook on Smart Technology for Aging, Disability and Independence*. John Wiley & Sons.

Broadbent, D.E. (1977). The hidden preattentive processes. *American Psychologist*, 32, 109-118.

Brochard, R., Drake, C., Botte, M-C., & McAdams, S. (1999). Perceptual organization of complex auditory sequences: Effect of number of simultaneous subsequences and frequency separation. *J of Experimental Psychology: Human Perception and Performance, 25*(6), 1742-1759.

Brown, L.M., Brewster, S.A., & Purchase, H.C. (2006). Multidimensional tactons for non-visual information display in mobile devices. *Proc Eigth Int'l Conf Human Computer Interaction with mobile devices and Service (MobileHCI '06)*, 231-238.

Chambrin, M.C., Ravaux, P., Calvelo-Aros, D., Jaborska, A., Chopin, C., and Boniface, B., (1999). Multicentric study of monitoring alarms in the adult intensive care unit (ICU): a descriptive analysis. *Intensive Care Med*, 25, 1360-1366.

Charabati, S., Bracco, D., Mathieu, P.A., and Hemmerling, T.M. (2009). Comparison of four different display designs of a novel anaesthetic monitoring system, the 'integrated monitor of anesthesia (IMA™)'. *British Journal of Anaesthesia, 103*(5), 670-677.

Cholewiak, R.W. & Collins, A.A. (1991). Sensory and physiological bases of touch. In Heller, M.A & Schiff, W. (Eds.), *The Psychology of Touch*. Hillsdale, NJ: Lawrence Erlbaum Associates. 23-60.

Coiera, E.W., Tombs, V.J. & Clutton-Brock, T.H. (1996). Attentional overload as a fundamental cause of human error in monitoring. Hewlett-Packard Laboratories Technical Report. Available from: http://www.coiera.com/papers/attention.pdf

Cook, R.I. & Woods, D.D. (1994). Operating at the sharp end: The complexity of human error. In Bogner, M.S. (Ed.), *Human Error in Medicine*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. 255-310.

Cook, R.I. & Woods, D.D. (1996). Adapting to new technology in the operating room. *Human Factors, 38*(4), 593-613.

Cooper, J.B., Newbower, R.S., Long, C.D., & McPeek, B. (1978). Preventable anesthesia mishaps: A study of human factors. *Anesthesiology, 49*. 399-406.

Craig, J.C. (1983). Some factors affecting tactile pattern recognition. *International Journal of Neuroscience*, 19, 47-58.

Dain, S. (2003). Current equipment alarm sounds: Friend or foe? *Canadian Journal of Anesthesiology, 50* (3), 209-214.

Drews, F.A. and Westenskow, D.R. (2006). The right picture is worth a thousand words: Data displays in anesthesia. *Human Factors, 48*(1), 59-71.

Drews, F.A., Syroid, N., Agutter, J., Strayer, D.L., and Westenskow, D.R. (2006). Drug delivery as control task: Improving performance in a common anesthetic task. *Human Factors, 48*(1), 85-94.

Evans, P.M. & Craig, J.C. (1986). Temporal integration and vibrotactile backward masking. *J. Experimental Psychology, 12*(2), 160-168.

Fitch, W.T. & Kramer, G. (1994). Sonifying the body electric: Superiority of an auditory over a visual display in a complex, multivariate system. In Kramer, G. (Ed.), *Auditory Display: Sonification, Audification, and Auditory Interfaces*. New York: Addison-Wesley.

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

Fried, R. & Imhoff, M. (2004). On the online detection of monotonic trends in time series. *Biometrical Journal, 46*(1), 90-102.

Gaba, D.M. & Lee, T. (1990). Measuring the workload of the anesthesiologist. *Anesthesia Analgesia, 71*, 354-361.

Gaba, D.M. (1994). Human error in dynamic medical domains. In Bogner, M.S. (Ed.), *Human Error in Medicine*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. 197-224.

Gallace, A., Tan, H.Z., & Spence, C. (2007). The body surface as a communication system: The state of the art after 50 years. *Presence, 16*(6), 655-676.

Geldard, F.A. (1960). Some neglected possibilities of communication. *Science, 131*, 1583-1588.

Görges, M., Markewitz, B.A., & Westenskow, D.R. (2009). Improving alarm performance in the medical intensive care unit using delays and clinical context. *Anesthesia & Analgesia, 108*(5), 1546-1552.

Görges, M. & Staggers, N. (2007). Evaluations of physiological monitoring displays: A systematic review. *Journal of Clinical Monitoring and Computing, 22*, 45-66.

Greenspan, J.D. & Bolanowski, S.J. (1996). The psychophysics of tactile perception and its peripheral physiological basis. In Kruger, L. (Ed) *Pain and Touch*. Elsevier.

Grundgeiger, T., Liu, D., Sanderson, P.M., Jenkins, S., & Leane, T. (2008). Effects of interruptions on prospective memory performance in anesthesiology. *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting*, 808-812.

Haimowitz, I.J., Le, P.P., Kohane, I.S. (1995). Clinical monitoring using regression-based trend templates. *Artificial Intelligence in Medicine, 7*, 473-496.

Haines , R. F. (1991). A breakdown in simultaneous information processing. In Obrecht, G. & Stark, L. W. (Eds.), *Presbyopia Research: From Molecular Biology to Visual Adaptation*. New York: Plenum. 171-175.

Hansson, R., Ljungstrand, P. and Redström, J. (2001). Subtle and public notification cues for mobile devices. In Abowd, G.D., Brumitt, B., and Shafer, S.A.N. (Eds), *Lecture Notes in Computer Science, Ubiquitous Computing, Vol. 2201*. Springer-Verlag Berlin Heidelberg. 240-246.

Health Grades, Inc. (2008). *The fifth annual HealthGrades patient safety in American hospitals study*. Available from: http://www.healthgrades.com/media/dms/pdf/patientsafetyinamericanhospitalsstudy2008.pdf

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

Imhoff, M. and Kuhls, S. (2006). Alarm algorithms in critical care monitoring. *Anesthesia and Analgesia*, 102, 1525-1537.

Ishii, H., Wisneski, C., Brave, S., Dahley, A., Gorbet, M., Ullmer, B., & Yarin, P. (1998). ambientROOM: Integrating ambient media with architectural space. *Proceedings of Computer-Human Interaction 1998 Conference Summary on Human Factors in Computing Systems*, 173-174.

James, R.H. (2003). 1000 anesthetic incidents: Experience to date. *Anaesthesia, 58*, 856-863.

Jones, L.A. & Sarter, N.B. (2008). Tactile displays: Guidance for their design and application. *Human Factors, 50*(1), 90-111.

Kesting, I.G., Miller, G.T., & Lockhart, C.H. (1988). Auditory alarms during anesthesia monitoring. *Anesthesiology, 69*, 106-107.

Kohn, L.T., Corrigan, J.M. and Donaldson M.S. (2000). *To err is human: Building a safer health system.* National Academy Press, Washington, DC.

Kramer, G. (1994). Some organizing principles for representing data with sound. In Kramer, G. (Ed.), *Auditory Display: Sonification, Audification, and Auditory Interfaces*. New York: Addison-Wesley. 79-94.

Lacherez, P., Seah, E.L., & Sanderson, P. (2007). Overlapping melodic alarms are almost indiscriminable. *Human Factors, 49*(4), 637-645.

Lawless, S.T. (1994). Crying wolf: False alarms in a pediatric intensive care unit. *Critical Care Med, 22*, 981-985.

Leape, L.L. (1994). Error in medicine. *JAMA, 272*(23). 1851-1857.

Lee, J.D., Hoffman, J. and Hayes, E.M. (2004). Collision warning design to mitigate driver distraction. *Proceedings of CHI 2004,* Vienna, Austria. 65-72.

Lee, J.D. & See, K.A. (2004). Trust in technology: Designing for appropriate reliance. *Human Factors, 46*(1), 50-80.

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 & Analgesia, 109*(4), 1135-1146.

Loeb, R.G. (1993). A measure of intraoperative attention to monitor displays. *Anesthesia & Analgesia, 76*, 337-341.

Loeb, R.G. (1994). Monitor surveillance and vigilance of anesthesia residents. *Anesthesiology, 80*, 527-533.

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

Mack, A. & Rock, I. (1998). *Inattentional Blindness*. Cambridge, MA: MIT Press.

MacLean, K.E. (2008a). Foundations of transparency in tactile information design. *IEEE Transactions on Haptics, 1*(2), 84-95.
MacLean, K.E. (2008b). Haptic interaction design for everyday interfaces. In Carswell, M. (Ed.) *Reviews of Human Factors and Ergonomics, Vol. 4.* Santa Monica, CA: Human Factors and Ergonomics Society. 149-194.
MacLean, K.E. (2009). Putting haptics into the ambience. *IEEE Transactions on Haptics, 2*(3), 123-135.
MacLean, K.E. & Enriquez, M. (2003). Perceptual design of haptic icons. *Proceedings of Eurohaptics.* 351- 363.
McGuirl, J.M. & Sarter, N.B. (2006). Supporting trust calibration and the effective use of decision aids by presenting dynamic system confidence information. *Human Factors, 48*(4), 656-665.
Meredith, C. & Edworthy, J. (1995). Are there too many alarms in the intensive care unit? An overview of the problems. *Journal of Advanced Nursing, 21*, 15-20.
Meyer, J. & Bitan, Y. (2002). Why better operators receive worse warnings. *Human Factors, 44*(3), 343-353.
Michels, P., Gravenstein, D., and Westenskow, D.R. (1997). An integrated graphic data display improves detection and identification of critical events during anesthesia. *J of Clinical Monitoring, 13*, 249-259.
Momtahan, K., Hetu, R. and Tansley, B. (1993). Audibility and identification of auditory alarms in the operating room and intensive care unit. *Ergonomics*, 36, 1159-1176.
Mosier, K.L. (1997). Myths of expert decision making and automated decision aids. In Zsambok, C. & Klein, G. (Eds.), *Naturalistic Decision Making*. Hillsdale, NJ: Erlbaum.
Ng, G., Barralon, P., Schwarz, S.K.W., Dumont, G., & Ansermino, J.M. (2008). Evaluation of a tactile display around the waist for physiological monitoring under different clinical workload conditions. *Proceedings of 30th Annual International IEEE EMBS Conference*, 1288-1291.
Ng, J.Y.C. & Man, J.C.F. (2004). Vibro-monitor: A vibrotactile display for physiological data monitoring. *Human Interface Technologies Conference, University of British Columbia*, 1-8.
Ng, J.Y.C., Man, J.C.F., Fels, S., Dumont, G., & Ansermino, J.M. (2005). An evaluation of a vibro-tactile display prototype for physiological monitoring. *Anesthesia Analgesia, 101*, 1719-1724.
Norman, D.A. (2002). *The Design of Everyday Things*. New York: Basic Books.
Parasuraman, R. & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*(2), 230-253.
Pasquero, J. (2006). Survey on communication through touch. Technical Report: TR-CIM 06.04. Center for Intelligent Machines, Department of Electrical and Computer Engineering, McGill University. August.
Pomerantz, J.R. and Pristach, E.A. (1989) Emergent features, attention, and perceptual glue in visual form perception. *Journal of Experimental Psychology: Human Perception & Performance, 15*, 635-649.
Posner, M.I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32, 3-25.
Rady, M.Y., Ryan, T., & Starr, N.J. (1998). Perioperative determinants of morbidity and mortality in elderly patients undergoing cardiac surgery. *Critical Care Medicine, 26*, 225-235.
Remington, R.W., Johnston, J.C., & Yantis, S. (1992). Involuntary attentional capture by abrupt onsets. *Perception and Psychophysics, 51*, 279-290.

Rensink, R.A., O'Regan, J.K., & Clark, J.J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8*(5), 368-373.

Roizen, M.F. & Fleisher, L.A. (2002). Abdominal aortic aneurysm repair. In Roizen, M.F. & Fleisher, L.A. (Eds.), *Essence of Anesthesia Practice, 2nd Edition*. Philadelphia: Saunders. 351.

Ross, B., Ormerod, D.F., Hyde, J.P., & Fine, M. (2002). Use of a head-mounted display of patient monitoring data to enhance anesthesiologists' response to abnormal clinical events. *Society for Technology in Anesthesia Annual Meeting (STA 2002)*.

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

Sanderson, P.M., Watson, M.O., Russell, W.J., Jenkins, S., Liu, D., Green, N., Llewelyn, K. , Cole, P., Shek, V., & Krupenia, S.S. (2008). Advanced auditory displays and head-mounted displays: Advantages and disadvantages for monitoring by the distracted anesthesiologist. *Anesthesia Analgesia, 106*, 1787-1797.

Sanderson, P.M., Liu, D., & Jenkins, S.A. (2009). Auditory displays in anesthesiology. *Current Opinion in Anesthesiology, 22*, 788-795.

Sanderson, P.M., Watson, M.O., & Russell, W. (2005). Advanced patient monitoring displays: Tools for continuous informing. *Anesthesia Analgesia, 101*(1): 161-168.

Sarter, N.B. (2000). The need for multisensory interfaces in support of effective attention allocation in highly dynamic event-driven domains: The case of cockpit automation. *International Journal of Aviation Pschology, 10*(3), 231-245.

Sarter, N.B. (2002). Multimodal information presentation in support of human-automation communication and coordination. In: Salas, E. (Ed.), *Advances in human performance and cognitive engineering research*, JAI Press, New York, 13-36.

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

Sarter, N.B. & Schroeder, B. (2001). Supporting decision making and action selection under time pressure and uncertainty: The case of in-flight icing. *Human Factors, 43*(4), 573-583.

Schoenberg, R., Sands, D.Z., & Safran, C. (1999). Making ICU alarms meaningful: a comparison of traditional vs. trend-based algorithms. *Proceedings of AMIA Symposium,*379-383.

Seagull, F.J. and Sanderson, P.M. (2001). Anesthesia alarms in context: An observational study. *Human Factors*, 43 (1), 66-78.

Seagull, F.J., Wickens, C., & Loeb, R.G. (2001). When is less more? Attention and workload in auditory, visual, and redundant patient-monitoring conditions. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting*, Santa Monica, CA, 1395-1399.

Sherrick, C.E. & Cholewiak, R.W. (1986). Cutaneous sensitivity. In Boff, K., Kaufman, L., & Thomas, J. (Eds.), *Handbook of Perception and Human Performance, Vol 1: Sensory Processes and Perception*. John Wiley & Sons.

Sklar, A.E. & Sarter, N.B. (1999). Good vibrations: Tactile feedback in support of attention allocation and human-automation coordination in event-driven domains. *Human Factors, 41*(4), 543-552.

Sorkin, R. D., (1988). Why are people turning off our alarms? *Journal of the Acoustical Society of America, 84*, 1107-1108.

Sorkin, R. D., Kantowitz, B. H., & Kantowitz, S. C. (1988). Likelihood alarm displays. *Human Factors, 30*(4), 445-459.

Sorkin, R.D. & Woods, D.D. (1985). Systems with human monitors: A signal detection analysis. *Human-Computer Interaction, 1*, 49-75.

Spence, C., McDonald, J., & Driver, J. (2004). Exogenous spatial-cuing studies of human cross-modal attention and multisensory integration. In Spence, C. & Driver, J. (Eds.), *Crossmodal Space and Crossmodal Attention*. New York: Oxford University Press. 277-320.

Sukuvaara, T., Koski, E.M.J., Mäkivirta, A., & Kari, A. (1993). A knowledge-based alarm system for monitoring cardiac operated patients – technical construction and evaluation. *Int J Clin Monit Comput, 10*(2), 117-126.

Trafton, J.G., & Monk, C.A. (2008). Task interruptions. In Boehm-Davis, D.A. (Ed.), *Reviews of human factors and ergonomics, Vol. 3*. Santa Monica, CA: Human Factors and Ergonomics Society. 111-126.

Ulrich, R.S., Zimring, C., Zhu, X., Dubose, J., Seo, G-B., Choi, Y-S., Quan, X., & Joseph, A. (2008). A review of the research literature on evidence-based healthcare design (Part 1). *Health Environments Research & Design, 1*(3), 61-125.

Via, D.K., Kyle, R.R., Kaye, R.D., Shields, C.H., Dymond, M.J., Damiano, L.A., & Mongan, P.D. (2003). A head mounted display of anesthesia monitoring data improves time to recognition of crisis events in simulated crisis scenarios. *Society for Technology in Anesthesia Annual Meeting (STA 2003)*.

Wachter, S.B., Johnson, K., Albert, R., Syroid, N., Drews, F., and Westenskow, D. (2006). The evaluation of a pulmonary display to detect adverse respiratory events using high resolution human simulator. *Journal of the American Medical Informatics Association, 13*(6), 635-642.

Walsh, T. & Beatty, P.C.W. (2002). Human factors error and patient monitoring. *Physiological Measurement, 23*, R111-R132.

Wallace, M.S., Ashman, M.N., & Matjasko, M.J. (1994). Hearing acuity of anesthesiologists and alarm detection. *Anesthesiology, 81*, 13-28.

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

Watson, M.O., Sanderson, P.M., & Russel, W.J. (2004). Tailoring reveals information requirements: The case of anesthesia alarms. *Interactive Computing, 16*, 271-293.

Webb, R.K., van der Walt, J.H., Runciman, W.B., Williamson, J.A., Cockings, J., Russell, W.J., and Helps, S. (1993). Which monitor? An analysis of 2000 incident reports. *Anaesthesia and Intensive Care, 21*, 529-542.

Welch, R.B. & Warren, D.H. (1986). Intersensory interactions. In Boff, K., Kaufman, L., & Thomas, J. (Eds.), *Handbook of Perception and Human Performance, Vol 1: Sensory Processes and Perception*. John Wiley & Sons.

Wickens, C.D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Sciences, 3*, 159-177.

Wickens, C.D. (2008). Multiple resources and mental workload. *Human Factors, 50*(3), 449-455.

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.

Wisneski, C., Ishii, H., Dahley, A., Gorbet, M., Brave, S., Ullmer, B., & Yarin, P. (1998). Ambient displays: Turning architectural space into an interface between people and digital information. *Proceedings of the International Workshop on Cooperative Buildings (CoBuild '98)*, 22-32.

Woods, D.D. (1995). The alarm problem and directed attention in dynamic fault management. *Ergonomics*, 38 (11), 2371-2393.
Woods, D.D., O'Brien, J., & Hanes, L.F. (1987). Human factors challenges in process control, the case of nuclear power plants. In Salvendy, G. (Ed.), *Handbook of Human Factors/Ergonomics*. Wiley, New York.
Xiao, Y., Mackenzie, C.F., Seagull, F.J., & Jaberi, M. (2000). Managing the monitors: An analysis of alarm silencing activities during an anesthetic procedure. *Proceedings of the IEA/HFES 2000 Congress*, 4250-4253.
Xiao, Y. & Seagull, F.J. (1999). An analysis of problems with auditory alarms: Defining the roles of alarms in process monitoring tasks. *Proceedings of the HFES 43rd Annual Meeting*, 256-260.
Zhang, Y., Drews, F.A., Westenskow, D.R., Foresti, S., Agutter, J., Bermudez, J.C., Blike, G., and Loeb, R. (2002). Effects of integrated graphical displays on situation awareness in anaesthesiology. *Cognition, Technology & Work, 4*, 82-90.

Chapter 2

Two Important Considerations in Tactile Interface Design: Tactile Change Blindness and Processing Code Interference

A variety of data-rich complex domains stand to benefit from the introduction of tactile displays that offload the visual and auditory information channels. Yet, a better understanding of factors that affect tactile information processing is needed to ensure the robustness of these interfaces. The first part of the following chapter (Part A) describes how a tactile analog of visual "change blindness" can interfere with processing tactile signals at the perceptual processing stage. Recent research has demonstrated tactile change blindness effects for one important dimension of tactile processing: the number of body locations where vibrations were presented. The first study described in this chapter investigated these effects within another important dimension for tactification design: vibration intensity. Additionally, because the tactification was designed for use in a complex multitask environment, this study investigated whether the addition of a secondary task exacerbates the change blindness effects. The findings demonstrate particular patterns which are prone, and others which are relatively resilient, to the effects of change blindness in detection of vibration intensity. These findings were then used as guidelines for tactification design.

The second half of this chapter concerns interference that can occur at later processing stages, "downstream" of perception. Past research demonstrating the benefits of tactile and multimodal interface design has primarily focused on the employment of separate presentation modalities in the effectiveness of information processing. However, it is not well known to what extent the interpretation of tactile patterns is affected by another attribute of information: the information processing codes of concurrent tasks. Two studies set in a driving simulation specifically manipulated this dimension in a visual-tactile multitask set. The findings demonstrate the importance of considering cognitive processing code when designing tactile patterns which are to be interpreted in parallel with concurrent tasks.

Part A:

Tactile “Change Blindness” in the Detection of Vibration Intensity

INTRODUCTION

As with visual and auditory displays, the design of effective tactile displays requires careful consideration of a variety of factors. For example, the ease of interpreting tactile communications can depend heavily on how naturally the signal maps to the encoded message (MacLean, 2008), and on the degree to which cognitive processing of the tactile signal interferes with processing demands associated with other ongoing tasks (e.g., Ferris & Sarter, 2010). Additionally, performance-shaping perceptual and attentional phenomena that had first been demonstrated for the visual and auditory channels appear to affect tactile information processing as well. These phenomena include vibrotactile masking (e.g., Evans & Craig, 1986; Tan, Reed, Delhorne, Durlach, & Wan, 2003), crossmodal orientation of spatial attention (e.g., Ho, Tan, & Spence, 2005), and, possibly, change blindness.

Change blindness refers to the fact that humans are surprisingly poor at detecting large changes in a visual scene when the changes coincide with the presentation of a masking or distracting (visual) stimulus which is sometimes referred to as a “transient” (e.g., Levin & Simons, 1997; O’Regan, Resink, & Clark, 1999). Evidence has emerged recently of this phenomenon occurring in the tactile modality. In particular, research has shown a decreased ability to detect changes in the number of body locations where vibrations were presented when different types of transient stimuli or a blank interstimulus interval coincided with those changes (Gallace, Tan, & Spence, 2005; 2006; 2007). Presentation location on the body represents only one of several dimensions of tactile display; yet, to date, it is the only one that has been examined for change blindness effects. To ensure robustness of tactile information processing and presentation, it is necessary to investigate whether these effects also occur in other dimensions, such as vibration intensity. Additionally, since most applications of tactile displays involve interpreting tactile messages as part of a larger task set, it is important to understand how the effects of tactile change blindness are modulated by concurrent tasks that compete for attentional resources and impose a higher level of overall workload.

In the present study, participants conducted multiple tasks modeled after those of an anesthesiologist in a simulated hospital operating room. One task involved interpreting a tactile display which communicated changes in a simulated patient's blood pressure through corresponding changes in the intensity of a vibration pattern. At times, "noise" in the display system was simulated and created four different types of transient effects which sometimes co-occurred with an intensity change: a blank interstimulus interval, two different types of masking stimuli, and a gradual transition. Participants had to determine and indicate whether or not an intensity change had occurred, both in isolation and while completing a demanding secondary task. Expectations were that participants' ability to accurately detect changes in vibrotactile intensity would be affected by 1) the magnitude of the intensity change (larger changes would be more reliably detected), 2) the presentation condition (change detection for presentations in which each of the four transient effects were present would be worse than in a baseline "no transient" condition), and 3) the requirement to conduct a concurrent secondary task (change detection in dual-task conditions would be worse). The study also explored which of the four types of transients were most detrimental to change detection (thus demonstrating the strongest vibrotactile intensity change blindness), and whether change detection was differentially affected in the individual presentation conditions by the demands of the secondary task. The findings from this study will add to the knowledge base in tactile perception and attention and can be used to inform the design of vibrotactile displays for a wide range of work environments.

METHOD

Twenty-five students from the University of Michigan participated in this study. Data from two participants were removed from analysis due to an apparent misunderstanding of instructions which resulted in incomplete datasets (N = 23; 15 males and 8 females; mean age = 22.4, stdev = 2.0). Each participant played the role of an anesthesiologist and was responsible for two tasks in a simulated operating room. The primary task involved adjusting the delivery rate of a drug infusion to a simulated patient in response to changes in the patient's blood pressure, which were relayed via patterns of vibrations applied to their forearm. Each vibration pattern presentation constituted an experimental trial. The secondary task involved performing an endotracheal intubation (insertion of a breathing tube into the airway of the anesthetized patient). Half of the trials required participants to perform only the primary task (isolation case), and the other half required both tasks to be conducted concurrently (dual-task case). Figure 2.1 shows the layout for the visual displays associated with each task.

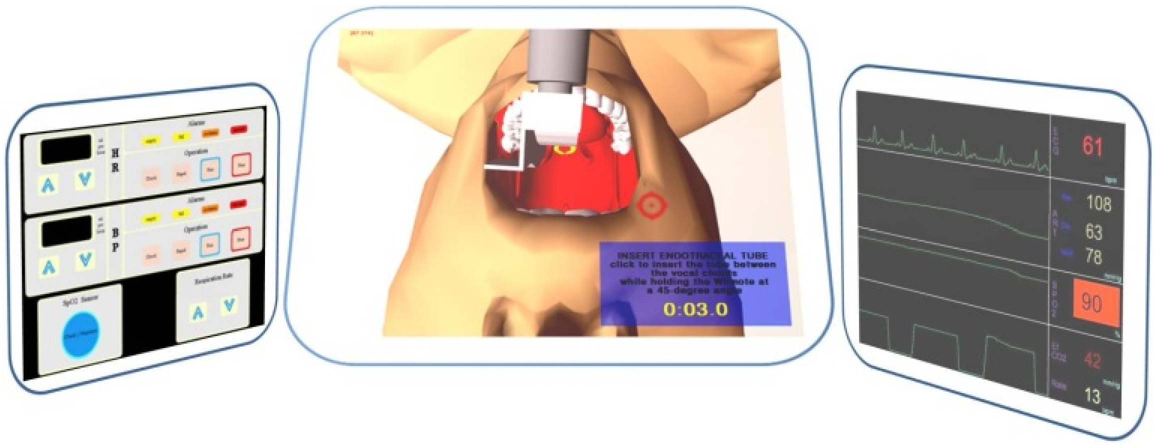


Figure 2.1: The simulation setup, distributed across three separate desktop monitors. Left to right: drug infusion controls (touch screen), the intubation simulation display, and the physiological data display.

The tactile display consisted of four vibrating “tactor” devices (C-2 tactors developed by Engineering Acoustics, Inc.) secured to the participants’ non-dominant forearm with an elastic compression sleeve. The forearm was chosen after pilot testing of multiple body sites showed superior performance in detecting the vibration intensity change patterns used in this study. One tactor was secured to each of the dorsal and palmar surfaces of the wrist (top and bottom, respectively, when the hand is palm-down) and to similar positions near the elbow. Five-second vibrations were presented via a circular “skin contactor” with diameter 7.5 mm oscillating at 250 Hz with a maximum displacement of 1 mm. This resulted in vibration intensities on the order of 20 dB above the sensory threshold. Noise-cancelling headphones playing ambient noise recorded in a hospital operating room and simulation sounds (such as the auditory pulse oximetry sonification) masked the sound of tactor activation. This assured that detecting changes in the signal could only be done via tactile (and not auditory) perception.

Primary Task: Response to Tactile Displays

For the purposes of this study, participants were told that an experimental tactile display was being evaluated for its ability to relay data about longer-term levels and trends in blood pressure. This data involved complex filtering of raw blood pressure data over an extended time period and therefore would not necessarily directly map to the real-time blood pressure data which was presented visually on the patient's complete physiological data display (right side of Figure 2.1; this display was included for scenario realism).

For each trial, participants were instructed to indicate whether or not they detected a change in the intensity of the vibration during its 5-second presentation, i.e., if the intensity at the beginning of the presentation was different from the intensity at its conclusion. Participants were told such changes were related to long-term trends in blood pressure (increasing intensity for increasing pressure). Half of the trials involved such a change ("change" trials) while the other half presented a constant intensity ("no change" trials). The possible intensity levels were approximately 14, 17, and 20 dB above sensory threshold, thus possible changes included increasing or decreasing either 3 dB or 6 dB in magnitude. Pilot tests verified that changes involving each intensity pairing (6 possible changes between the 3 intensities) were clearly perceptible in unmasked step change presentations (the "Step" presentation condition illustrated in Figure 2.2). To minimize the effects of vibrotactile adaptation – the decreased sensitivity to vibrotactile stimuli that results from prolonged vibration exposure at a single location (e.g., Verillo & Gescheider, 1977) – the presentation location was randomized for each trial. This spatial unpredictability also made it more difficult to compare stimulus intensities *between* trials (because they were often at separate locations). This minimized any advantage that may have been gained by noticing that a new trial began with a higher or lower intensity than the previous trial, which may provide a clue as to the absolute intensity level and allow participants to focus on detecting changes in just one direction. For example, knowing that a trial begins at a lower intensity would allow participants to focus on detecting an increase in intensity, rather than a change which could be an increase or decrease. Participants were told that their tasks required only the identification of intensity changes; the specific location where vibrations were presented communicated a measure of hemodynamics that did not need to be considered for the experimental task set.

Each vibration presentation could be categorized as one of five presentation conditions:

Step, Blank Interval, Masked Interval, Mudsplash, and Gradual (see Figure 2.2). These conditions represent different paradigms and “transients” used in visual change blindness studies. Except for the Gradual condition (which was modeled after fade in/fade out changes which can be very difficult to detect in a visual scene, e.g., Simons, Franconeri, & Reimer, 2000), similar versions of each of the conditions have also been investigated in studies of tactile change blindness for which participants had to detect the number of locations where vibrations were presented (e.g., Gallace et al., 2005; 2006; 2007). In this study, participants were told that the off-time in the Blank Interval condition and the transient stimuli (vibration presentations that coincided with the time when a change could occur in the information-carrying stimulus) in the Masked Interval and Mudsplash conditions were due to “bugs” or “noise” in the display technology.

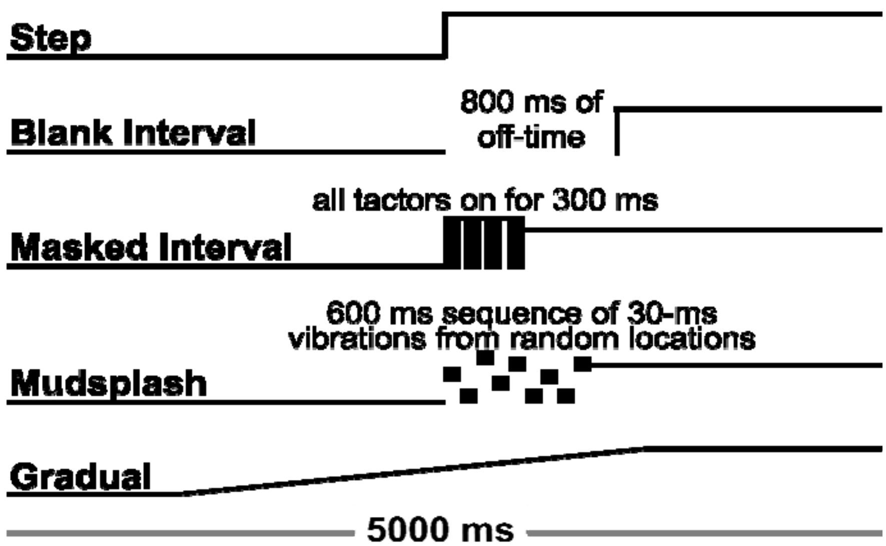


Figure 2.2: Intensity changes for the five presentation conditions.

The temporal characteristics of each condition were designed to be similar to those of the previous tactile change blindness studies. The durations of vibration presentation on either side of the change minimized the effects of sequential vibrotactile masking (Evans & Craig, 1986; Tan et al., 2003) to assure that the vibrations before and after transient stimuli could be clearly perceived (masking only the change in the signal). Pilot testing was used to refine the signals to most strongly elicit the desired effects. Except for Gradual presentations, each condition first involved a vibration for 2000 to 2500 ms (randomly distributed) at the first signal intensity. This vibration was followed by a transient stimulus or the absence of stimulation for up to 800 ms.

Finally, the remaining duration of the 5000-ms presentation would be at a second intensity (which may or may not be the same as the first). Gradual presentations involved approximately 1000 ms at the first intensity, then a linear transition to the second intensity over 2500 ms, and finally 1500 ms at the second intensity.

The five types of vibration presentation conditions were balanced across 54 experimental trials and presented in a random order. Each of the first four presentation conditions was represented in 12 trials: 6 “no change” trials and 6 “change” trials in which each of the possible changes between the 3 intensities were represented. The Gradual presentation condition had only the 6 change trials, since by definition this condition required a change to occur (i.e., a gradual transition could not occur without a change in intensity). In the isolation case, experimenters initiated each trial and participants noted whether they felt the signal intensity increase, decrease, or remain constant for the duration of the trial. In the dual-task case, trials occurred automatically at randomized intervals between 20 and 30 seconds apart (mean inter-trial interval was about 23 seconds). When changes in intensity were detected, participants were instructed to increase or decrease the dosage rate on a simulated (touchscreen) blood pressure drug infusion pump to counteract the perceived pressure change. A lack of response for 10 seconds after each presentation was recorded as a “no change” response.

Secondary Task: Intubate Simulated Patient

The process of endotracheal intubation involves a series of activities which can take several minutes and requires extended periods of focused visual attention and actions which must be performed with both hands. The simulated intubation task in this study involved a “worst case scenario” which required several unusual steps, such as suctioning vomit from the airway. A mouse and wireless gyroscopic controller were used to interact with the patient and various intubation tools. Participants followed the instructions on the screen to complete all of the steps in the intubation. More details on the specifics of this simulation can be found in Chapter 4.

Participants were instructed that their top priority was to respond to the vibration presentations, but to finish the intubation task as quickly as possible. The end of this task involved a series of patient charting steps that continued until all of the experimental trials had been completed, which took approximately 10 minutes. An informal “top score” in the number of charting steps completed before the 10 minute time limit for the intubation task served to motivate appropriate levels of attention to this task.

Experimental Design

The primary independent variables in this study were presentation condition (5 conditions), whether a trial involved an intensity change or not (change, no change), the magnitude of the change (3 dB or 6 dB) and the number of tasks involved (isolation case, dual-task case). Responses to the trials were coded as either “no change” or “change”, and performance for each participant was tallied in a Signal Detection Theory (SDT) framework: hits, misses, correct rejections, and false alarms. Measures of sensitivity (d’) and response bias (C) were then calculated for each participant under each combination of factor levels and compared. Because Gradual presentations did not include “no change” trials (thus d’ and C could not be calculated), overall hit rates were also compared.

Procedure

Experimenters first described the “experimental tactile display” and familiarized participants with the simulation and required actions with each controller. Next, participants completed a 30-minute training session which allowed them to practice the intubation task and feel examples of each of the five presentation conditions, as well as each of the six gain changes (14 dB to 17 dB, 17 to 20, 14 to 20, 20 to 17, 20 to 14, and 17 to 14) in the Step presentation condition. At the end of the training session, participants were required to achieve 100% accuracy in determining whether an intensity change occurred or not in a set of Step presentations that included presentations at each of the four locations and each of the six gain changes (the order of locations and gain changes was randomized). Upon successful completion of the training session, participants completed the full set of 54 experimental trials in the isolation case and in the dual-task case. The order of the cases was balanced between subjects.

Data Analysis

For tactile change detection performance, repeated measures linear models (using the General Linear Model formulation in SPSS 16.0) were used to identify main effects, and two-tailed Fisher's LSD post-hoc tests were used to determine differences between means for significant effects.

SDT measures of sensitivity (d') and response bias (C) involve calculations with the z-scores of hit and false alarm rates, which approach +/- infinity when the rates are 100% or 0% (which they often were in this study). Therefore a standard correction was applied (Stanislav & Todorov, 1999) such that the maximum value for these rates was (2N-1)/(2N) and minimum rates were 1/(2N), where N is the number of trials over which each rate was calculated (N = 6 in most cases). Since the Gradual condition did not include "no change" trials, it was not considered in the SDT analysis.

RESULTS

Hit Rates

Hit rate, defined as the percentage of "change" trials which were correctly identified as involving changes in intensity, was found to be significantly affected by presentation condition ($F(4,19) = 16.944$; $p < .001$), magnitude of change ($F(1,22) = 62.242$; $p < .001$), and an interaction between the number of tasks and presentation condition ($F(4,19) = 5.920$; $p = .003$). Post-hoc tests showed that hit rates for 6 dB changes (mean hit rate: 76.1%) were considerably higher than those for 3 dB changes (52.5%). Comparing presentation conditions, hit rates for the Step (mean hit rate: 72.5%) and Blank Interval (72.5%) conditions did not differ from each other, but each differed significantly from hit rates in the other three conditions ($p < .005$ in all comparisons). The lowest hit rates were found in the Gradual condition (44.9%), which showed significantly lower hit rates than in the Masked Interval (58.3%; $p = .009$) and Mudsplash (56.9%; $p = .005$) conditions. The Masked Interval and Mudsplash conditions did not result in different hit rates.

An examination of the significant interaction shows that hit rates did not differ between the isolation and dual-task cases for any of the presentation conditions except for the Mudsplash condition, which showed hit rates in the isolation case (64.5%) to be significantly better than in the dual-task case (49.3%; $p = .003$). Hit rates are shown in Figure 2.3 for each presentation condition, at each magnitude of change, and with each number of tasks.

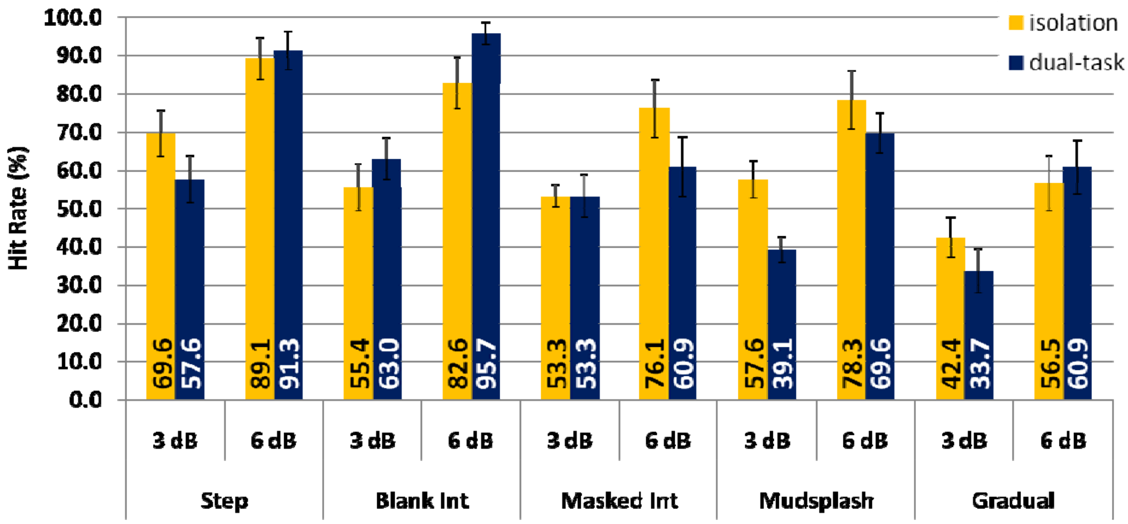


Figure 2.3: Hit rate for each presentation condition and each magnitude of change in the isolation and dual-task cases. Error bars represent standard error.

Sensitivity (d') and Response Bias (C)

Figure 2.4 shows the values of d' (bars associated with left axis) and C (diamonds associated with right axis) for each presentation condition in the isolation and dual-task cases. The sensitivity (d') of participants in detecting signal changes, which takes into account both hit and false alarm rates, significantly differed between presentation conditions ($F(3,20) = 22.321$; $p < .001$) and an interaction effect was also found between number of tasks and presentation condition ($F(3,20) = 5.486$; $p = .006$). Post-hoc tests showed that sensitivities for the Step (mean d' = 1.611) and Blank Interval (1.466) conditions were not statistically different, but sensitivities in those two conditions were both significantly higher than those in the Masked Interval (0.657) and Mudsplash (0.725) conditions ($p < .001$ in all comparisons). Sensitivities in the Masked Interval and Mudsplash conditions were not significantly different. Post-hoc analysis of the interaction effect showed that the Mudsplash condition was the only presentation condition that showed a difference in sensitivity between the isolation case (d' = 1.120) and dual-task case (0.331; $p = .001$).

The measure “C” relates the degree to which a participant is generally biased toward giving “change” (negative C values) or “no change” (positive C values) responses. While none of the experimental factors significantly affected the value of C, it is worth noting that values were above zero for every combination of factor levels (grand mean C = 0.136).

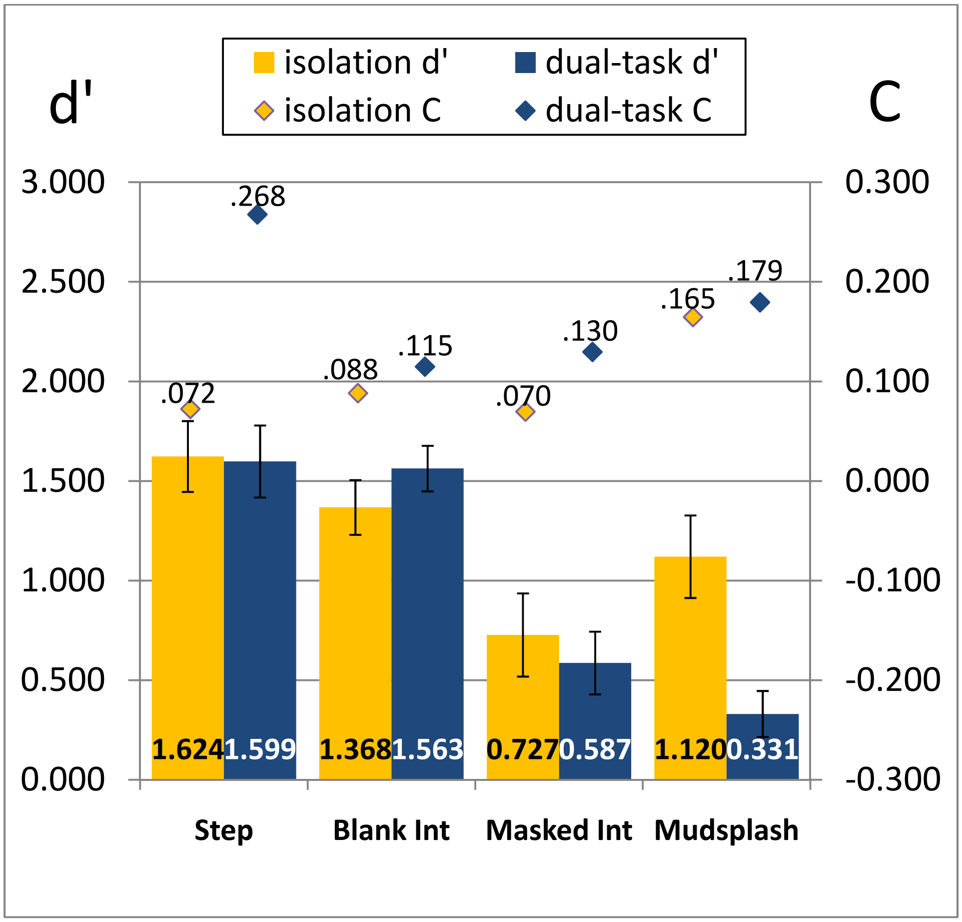


Figure 2.4: SDT analysis results. Sensitivity (d’): bars associated with the left axis; and response bias (C): diamonds associated with the right axis; for each presentation condition in the isolation and dual-task cases. Error bars represent standard error.

Intubation task

Each participant completed the secondary intubation task on time (within 10 minutes), and the number of charting tasks completed at the end was fairly consistent. No significant correlation was found between performance in change detection and intubation completion time.

DISCUSSION

The introduction of tactile displays in many data-rich environments, such as flight decks, car cockpits or the hospital operating room, can benefit operator performance by improving timesharing and offloading visual and auditory channels. However, ensuring the best performance with these displays requires that several factors be carefully considered in their design. These factors include perceptual and attentional phenomena that occur within the tactile channel, such as recently discovered tactile forms of change blindness. Previous studies have demonstrated how the ability to detect changes in one dimension of tactile display – the number of body locations where vibrations are applied – can be drastically reduced when the change occurs during a blank interstimulus interval or when it coincides with a presentation of vibrotactile masking stimuli (Gallace et al., 2005; 2006; 2007). The goal of this study was to investigate whether this effect extended to another dimension of a tactile signal which is prominent in tactile display design: vibration intensity.

The findings of the present study show that, not surprisingly, the ability to identify changes in vibrotactile intensity was better for changes that were larger in magnitude (6 dB vs. 3 dB). They also showed that the presentation condition clearly affected change detection performance, independent of the magnitude of change. Equivalently high change detection performances were found in the Step condition (no blank interval or transient stimuli coinciding with the intensity change) and in the Blank Interval condition; these two conditions showed better change detection performance than all other conditions. The lack of difference in performance in detecting intensity changes between the Step and Blank Interval conditions contrasts the findings of other tactile change blindness studies (Gallace et al., 2005; 2006). These studies showed that a blank interval coinciding with a change in the number of body locations where tactile stimuli were applied was sufficient to significantly impair detection of the change. The findings of the study described in this chapter suggest that the dimension of vibration intensity may be less susceptible than the dimension of vibration location to change blindness effects triggered by a blank interval between changes in the signal.

The Masked Interval and Mudsplash conditions – each of which involved the presentation of transient stimuli during the change in the underlying signal – did not differ from each other but each showed poorer change detection than the Step and Blank Interval conditions. These findings are similar to those found in previous studies involving the dimension of vibration presentation location (Gallace et al., 2006; 2007). They suggest that the vibrotactile masking effect may play a significant role in the occurrence of tactile change blindness, especially when the time interval between the onset of the masker (transient stimuli) and onset of information-carrying signal is very short, which is when the masking effect is strongest (Evans & Craig, 1986; Sherrick & Cholewiak, 1986; Tan et al., 2003). For longer intervals, masking effects may also interact with limitations in the ability to hold a detailed representation (including the level of intensity) of a tactile signal in short-term memory (Gallace, Tan, Haggard, & Spence, 2008).

Hit rate in the detection of intensity changes was worst overall in the Gradual presentation condition, which was modeled after a change blindness paradigm from the visual literature (Simons et al., 2000) that had not previously been investigated in the context of tactile change blindness. The mechanism affecting change detection in this condition may be vibrotactile adaptation, which decreases the sensitivity of mechanoreceptors to a persistent vibration presentation to an extent that depends on both the duration of the presentation and its intensity (Sherrick & Cholewiak, 1986; Verillo & Gescheider, 1977). This effect plays a significant role in the fundamental ability to detect changes in intensity, which is better with transitions that are larger in magnitude and/or shorter in duration (Goble & Hollins, 1993). This implies that if a tactile display uses intensity to communicate a continuous variable, it may be better to discretize the variable and represent changes in intensity in steps that are large enough to support reliable detection.

When discussing of the roles of vibrotactile masking and adaptation effects as contributing components of the tactile change blindness effect, it should be noted that the occurrence and strength of these effects depend on the mechanoreceptors involved in sensation of the vibratory stimulus. The 250 Hz signals used in this study, and the 290 Hz signals used in previous studies (Gallace et al., 2005; 2006; 2007), are within the range of maximum sensitivity for the Pacinian corpuscle (Cholewiak & Collins, 1991; Sherrick & Cholewiak, 1986). This mechanoreceptor is classified as “rapidly-adapting”, hence is more susceptible to adaptation (and masking) effects than other “slowly-adapting” receptors (such as SA Type I/Ruffini cylinders or SA Type II/Merkel disks) which may be more involved in perception of vibration at lower frequencies (Cholewiak & Collins, 1991; Sherrick & Cholewiak, 1986). It is reasonable to assume, therefore, that intensity changes at lower operating frequencies may be more resilient to tactile change blindness, though at the cost of decreased overall sensitivity to vibrations applied to hairy skin (Sherrick & Cholewiak, 1986).

One unexpected but interesting finding was that the addition of a secondary concurrent task did not affect change detection performance overall. However, significant interaction effects showed that one presentation condition, the Mudsplash condition, did show change detection performance to be significantly poorer when participants were concurrently conducting the intubation task, compared to in isolation. This may be explained by the relatively higher degree of focused attention required to "tune out" the mudsplash transient, which was a random sequence of vibrations over an interval about twice as long as the simpler Masked Interval transient. The ability to sustain this higher degree of focus may be hindered when some attentional resources are allocated to the concurrent intubation task. This finding suggests that extra caution should be given when designing tactile displays for attentionally-demanding environments which may be characterized by discrete environmental vibrotactile events which stimulate multiple body locations, (such as in a car, when driving over changing road surfaces). These patterns of vibrations may serve as a naturalistic version of the Mudsplash presentation condition, and can negatively impact the ability to interpret the tactile display, especially in attentionally-demanding situations. Similarly, the design of complex tactile displays that include multiple tactile presentation streams should consider the potential effects of change blindness. Even if the streams are presented to separate body locations, it may not be wise to rely solely on the dimensions of vibrotactile intensity or presentation location to communicate critical data, since changes in these dimensions may be surprisingly difficult to detect when the streams temporally overlap.

Because the introduction of tactile displays may be most beneficial in environments that heavily load the visual and auditory channels, it may also be necessary to consider *crossmodal* versions of change blindness which may negatively affect tactile (and/or visual, auditory) perception. Some recent studies provide evidence that change blindness may be multisensory in nature, at least within the spatial dimension (Auvray, Gallace, Tan, & Spence, 2007; Gallace & Spence, 2008). These studies showed, for example, that detection of changes in the body locations where either vibrations or visual stimuli were presented was impaired by a transient masking stimulus that could be either visual or tactile. If the same mechanism modulates the change blindness effects for vibrotactile intensity, these effects may be more likely to be elicited in complex multimodal environments such as the hospital operating room. Further, the vibrotactile presentations may affect perception of other critical visual and auditory stimuli.

Taken together, these findings suggest that designers of tactile and multimodal displays in these environments need to consider the effects of change blindness to ensure that critical information is reliably perceived. Designers may attempt to minimize the associated performance costs, for example, by using redundant signal dimensions and/or modalities to display critical data.

The results of this study complement recent studies by demonstrating a tactile analog of visual change blindness within the dimension of vibration intensity. They illustrate the extent to which change blindness can influence detection performance for vibrotactile intensity changes and thus highlight the need to consider this perceptual phenomenon in tactile display design. At the conclusion of this chapter (after Part B), the specific contributions to the tactification designs are briefly discussed.

REFERENCES

Auvray, M., Gallace, A., Tan, H.Z., & Spence, C. (2007). Crossmodal change blindness between vision and touch. *Acta Psychologica*, 126, 79-97.

Cholewiak, R.W. & Collins, A.A. (1991). Sensory and physiological bases of touch. In Heller, M.A. & Schiff, W. (Eds.), *The Psychology of Touch*. Lawrence Erlbaum Associates, Hillsdale, NJ. 23-60.

Evans, P.M. & Craig, J.C. (1986). Temporal integration and vibrotactile backward masking. *J. Experimental Psychology,* 12(2), 160-168.

Ferris, T., & Sarter, N. (2010). When content matters: The role of processing code in tactile display design. *IEEE Transactions on Haptics.*

Gallace, A. & Spence, C. (2008). The cognitive and neural correlates of “tactile consciousness”: A multisensory perspective. *Consciousness and Cognition, 17*, 370-407.

Gallace, A., Tan, H.Z., Haggard, P., & Spence, C. (2008). Short term memory for tactile stimuli. *Brain Research*, 1190, 132-142.

Gallace, A., Tan, H.Z., & Spence, C. (2005). Tactile change detection. *First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC’05)*, 12-16.

Gallace, A., Tan, H.Z., & Spence, C. (2006). The failure to detect tactile change: A tactile analogue of visual change blindness. *Psychonomic Bulletin & Review*, 13(2), 300-303.

Gallace, A., Tan, H.Z., & Spence, C. (2007). Do “mudsplashes” induce tactile change blindness? *Perception & Psychophysics*, 69(4), 477-486.

Goble, A.K. & Hollands, M. (1993). Vibrotactile adaptation enhances amplitude discrimination. *J. Acoustical Society of America*, 93(1), 418-424.

Ho, C., Tan, H.Z. & Spence, C. (2005). Using spatial vibrotactile cues to direct visual attention in driving scenes. *Transp. Res. Part F*, 8: 397-412.

Levin, D.T. & Simons, D.J. (1997). Failure to detect changes to attended objects in motion

pictures. *Psychonomic Bulletin Review*, 5, 644-649.
MacLean, K.E. (2008). Foundations of transparency in tactile information design. *IEEE Transactions on Haptics*, 1(2), 84-95.
O'Regan, J.K., Rensink, R.A., & Clark, J.J. (1999). Change-blindness as a result of 'mudsplashes'. *Nature, 398*, 34.
Sherrick, C.E. & Cholewiak, R.W. (1986). Cutaneous sensitivity. In Boff, K., Kaufman, L., & Thomas, J. (Eds.), *Handbook of Perception and Human Performance, Vol 1: Sensory Processes and Perception*. John Wiley & Sons.
Simons, D.J., Franconeri, S.L., & Reimer, R.L. (2000). Change blindness in the absence of a visual disruption. *Perception*, 29, 1143-1154.
Stanislav, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, and Computers*, 31, 137-149.
Tan, H.Z., Reed, C.M., Delhorne, L.A., Durlach, N.I., & Wan, N. (2003). Temporal masking of multidimensional tactual stimuli. *J Acoustical Society of America*, 114(6), 3295-3308.
Verillo, R.T. & Gescheider, G.A. (1977). Effect of prior stimulation on vibrotactile thresholds. *Sensory Processes, 1*, 292-300.

Part B:

The Role of Processing Code in Tactile Display Design

INTRODUCTION

The tactile channel combines a number of rather unique affordances. It is a proximal sense, in that presentation devices must be in contact with the skin. This characteristic allows for privatizing of displayed information. Touch is omnidirectional, such that, unlike vision, the perception of stimuli is not dependent on the spatial orientation of sensory organs. Touch is comparable to audition in its spatial discrimination capabilities and to vision in temporal discrimination (Geldard, 1960; Welch & Warren, 1986). These discrimination abilities, unlike for the other channels, deteriorate only minimally with age (Brewster, Wall, Brown, & Hoggan, 2008). Perhaps the most convincing argument for the development of tactile displays is that they help offload the visual and auditory channels, which tend to be utilized heavily in a number of data-rich domains. By shifting the display of some task-relevant information to the tactile channel, processing information for the overall task set can be more efficient, thus leading to improved multitasking. This follows from the assertion of Multiple Resource Theory (Wickens, 2002; 2008; Wickens & Hollands, 2000) that timesharing between multiple tasks results in minimal interference to the extent that the tasks require separate information processing resources, most notably different perceptual modalities.

Given the relatively narrow bandwidth of the tactile channel (Brown, Brewster, & Purchase, 2006; Tan, Durlach, Reed, & Rabinowitz, 1999), most applications of tactile displays have involved the communication of very simple, one-dimensional messages. For example, tactile cues have been used to provide a simple notification of system-initiated actions or system-detected events (Calhoun, Fontejon, Draper, Ruff, & Guilfoos, 2004; Sklar & Sarter, 1999), aid in spatial orientation in disorienting environments (e.g., show which way is “up”; Rupert, 2000; van Erp, Groen, Bos, & van Veen, 2002), and direct visual attention to displays or locations of interest (Ferris, Penfold, Hameed, & Sarter, 2006; Ho, Tan, & Spence, 2005; Tan, Gray, Young, & Traylor, 2003). There is, however, increasing interest in extending these displays to more complex signals that can communicate complete messages through iconic vibrotactile or haptic patterns (e.g., Brewster & Brown, 2004; Brown et al., 2006; MacLean & Enriquez, 2003). There are two schools of thought on appropriate ways to encode these messages (MacLean, 2008a; 2008b): 1) rely on representational or metaphorical symbolism in the signal modulation as it relates to the message; and 2) define abstract patterns by assigning meaning to modulations of various signal dimensions through somewhat arbitrary associations which must be learned.

If the signal itself serves as a metaphor for its meaning, minimal cognitive processing is required to decipher the message, and training time can also be minimized. For example, Chan, MacLean, & McGrenere (2005) used haptic renderings of a heart beating and a foot tapping to represent current control of a shared document and a request for access to the document, respectively. Learning time for this set of icons was short (3 minutes), and identification performance was consistently high, even under multitask conditions that imposed considerable workload. This type of iconic representation may be the best method for achieving "transparency" in tactile display (i.e., minimally loading the user's attentional resources; MacLean, 2008b), a quality that would make such displays much more attractive and usable for data-rich environments.

The downside to relying exclusively on metaphorically-derived tactile icons for display is a limited expressive range, driving researchers to explore the potential for abstract icons to express messages of greater complexity. Similar to the learning of abstract visual or auditory icons, longer training times are usually required before abstract tactile icons can be reliably deciphered, but with sufficient practice, information transmission rates can be surprisingly high (Craig, 1977; Gallace, Tan, & Spence, 2007a). For example, once properly trained in the abstract signals that defined a vibrotactile language called *Vibratese*, study participants were able to read 30 or more tactile words per minute (Geldard, 1960).

The extended practice required to achieve such high information transmission rates illustrates one of the primary challenges to the growth of tactile communication systems. Current tactile display technologies can provide a wide range of perceptual experiences to take full advantage of the perceptual discrimination capabilities of touch. However, tactile information processing may be less constrained by perceptual processing limitations than by a bottleneck that exists "downstream" in the later cognitive stages (Gallace et al., 2007a; MacLean, 2008a; 2008b). These later stages of processing may involve, for example, working memory operations such as comparing and combining components of a tactile pattern, identifying the pattern by matching the constructed mental representation to a template in memory, and interpreting the pattern by accessing stored information linked to the matched template. Supporting efficiency at both the perceptual and later cognitive stages is critical for the interpretability and practical usability of tactile and haptic icons.

Given a task that requires tactile/haptic icon interpretation, the introduction of a concurrent task set that engages different modalities (e.g., vision and/or audition) can be expected to lead to inefficiencies in the later cognitive processing stages and thus poorer overall performance if a) the overall task set imposes a combined information processing load that is sufficiently high, and/or b) the content of the information being processed in the concurrent tasks is sufficiently similar, such that there is competition for resources that are engaged for certain types of mental operations. With this in mind, it is somewhat surprising that few tactile/haptic icon identification studies to date have been conducted in realistic contexts or with concurrent task sets that are representative of a design environment (e.g., Brewster & King, 2005; Chan et al., 2005; Hameed, Ferris, Jayaraman, & Sarter, 2009; Pettitt, Redden, & Carstens, 2006). To help fill this gap, the studies described in this chapter required interpreting abstract tactile icons while performing concurrent visual tasks. In particular, the studies sought to determine how resource competition in one dimension of information processing – processing code (i.e., whether extracting information requires spatial or nonspatial mental operations) – ultimately affected multitask performance.

Multiple Resource Theory and the Role of Processing Code

Multiple Resource Theory (MRT; Wickens, 2002; 2008; Wickens & Hollands, 2000) represents one model of the structure of information processing resources that lends itself well to analyzing multitask performance when tasks engage multiple sensory modalities. Historically, MRT has been primarily employed to model visual and auditory task sets; the incorporation of the tactile channel remains one of the challenges/goals for its further development (Wickens, 2008). An adaptation of the most recent version of the model (e.g., Wickens, 2002; 2008) was developed to fit the visual/tactile task set of the studies described in this chapter and to emphasize the resource dimensions that are most relevant (see Figure 2.5). This adapted model illustrates the MRT assertion that multitask performance can be supported to the extent that concurrent tasks require different levels of three primary dimensions – processing stage (perception, cognition/working memory, and response), sensory modality (visual and tactile), and processing code (spatial and nonspatial) – and hence occupy different sections of the conceptual space (where solid lines denote divisions of resources) in Figure 2.5.

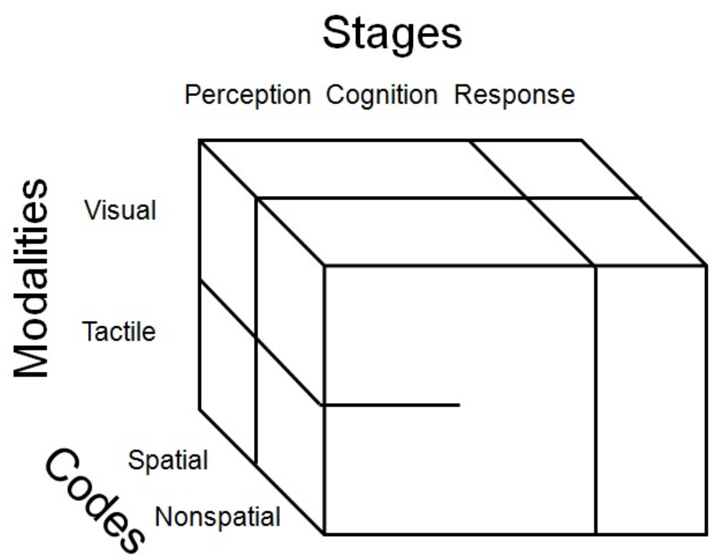


Figure 2.5: Adapted model of the structure of processing resources for a visual/tactile task set. Adapted from Wickens & Hollands, 2000 (see also Wickens, 2002; 2008).

According to this model, the interpretation of a tactile icon should minimally interfere with any concurrent visual (or auditory) tasks at the "perception" processing stage, because they occupy different levels of the modality dimension. For very simple tactile signals (e.g., attention-directing cues (Ferris, Penfold, Hameed, & Sarter, 2006; Ho et al., 2005) or simple notifications (Sklar & Sarter, 1999), this separation of modalities between tasks may be enough to ensure efficient performance. However, as tactile information becomes more complex, there is an increasing need to consider processing at the cognition stage, where the aforementioned working memory operations and icon identification occur. At this later stage, the ability to process concurrent streams of information should depend heavily on the processing codes of those streams.

For the purpose of this research, codes are described as either spatial, i.e., the information extracted from task-relevant stimuli relates to spatial relationships between stimulus components and/or the observer, or nonspatial, i.e., the extracted information has symbolic meaning or refers to identity within a category of information that is unrelated to spatial properties. For example, lane keeping in a vehicle primarily requires spatial processing while reading a speed limit sign primarily requires nonspatial processing. Using a map to locate a destination by its street address involves both codes.

Encoding Methods for Vibrotactile Icons

A number of different dimensions of a vibrotactile signal have been manipulated in the design of single- and multidimensional tactile icons, including signal frequency, intensity, waveform, rhythm, and stimulated body location. In this chapter, the methods employed to encode information in these icons will be described as they relate to either the spatial or nonspatial processing code. Generally, an icon that employs the dimension of body location to communicate (some of) its message is employing a spatial encoding method. In the simplest form, specific body locations can map to specific subsystems within a represented domain (e.g., Ford et al., 2008; Hameed et al., 2009), thus establishing context for the relayed message. Tactile devices can be arranged linearly, creating an axis to communicate additive information, such as time remaining until an appointment (Brown et al., 2006) or the level or direction of change of a system parameter [12]. Two- and three-dimensional (for example, with front/back as the third dimension in a torso display) arrays of devices on the body can also be used to present sequenced spatial patterns of vibrations to communicate rather complex information such as high-level navigation or action commands (Jones, Lockyer, & Piateski, 2006; Pettitt et al., 2006).

Nonspatial encoding methods for a tactile icon include modulating the operating frequency, amplitude/gain/intensity, waveform, or temporal properties of a vibration signal. There is a general consensus among tactile display researchers that signal dimensions of operating frequency and amplitude are not very desirable for encoding information (Brown et al., 2006; Geldard, 1960; Jones & Sarter, 2008; though see Hoggan & Brewster, 2007 for an example of their potential), primarily due to the apparent interaction between the two dimensions in perceived signal intensity (Verillo, Fraioli, & Smith, 1969) and the relatively low number of perceptually distinguishable levels for each dimension (van Erp, 2002). However, there may be promise in defining signal waveforms through changes in frequency and/or amplitude, and distinguishing signals by their different waveform patterns. For example, Brown et al. (2006) were able to create reliably distinguishable patterns of “roughness” with different frequencies of amplitude modulation.

A vibration signal can be described as a sequence of “pulses” (vibration presentations for predefined durations), and temporal variation of such pulses can be used as an effective nonspatial encoding method. Modulations include manipulating the duration of pulse on- or off-time, pulse repetition rate, number of pulses, or even creating complex pulse rhythms (Brewster & King, 2005; Brown et al., 2006; Ford et al., 2008; Hameed et al., 2009; Ternes & MacLean, 2008; van Erp & van Veen, 2004). For example, in a study by Hameed et al. (2009), the importance of attending to an interrupting task was communicated by the rate of pulses in the cue announcing the task.

A number of factors should be considered when deciding which encoding methods to employ for a new tactile display, including the constraints imposed by perceptual limits and device hardware, environmental conditions, and the comfort of the human receiving the message. If “transparency” in tactile communication is desired, the signal must be able to be learned and recognized in realistic conditions with minimal effort (MacLean, 2008b), promoting the use of encoding methods that maintain natural mappings between the message and the signal conveying it. Tactile communication transparency also requires that interpretation of the tactile signal be minimally disruptive of ongoing tasks, promoting encoding methods that support parallel processing of task-relevant information. It has been shown that competition for processing code resources is a major factor in the ability to support such parallel processing, especially with visual and auditory task sets. However, a better understanding of the role of processing code interference when processing complex tactile signals is needed and would allow a more informed approach in the design of tactile icons for multitask environments, such as aviation, medicine, or modern car cockpits.

Because of the considerable interest in tactile in-vehicle displays, and because of the significant safety benefits that may be gained from developing well integrated display systems into this domain, the present studies were conducted in a highly-controlled driving simulation. Simple visual tasks were used to emphasize either spatial or nonspatial information processing and were used to communicate lane navigation instructions to the driver. A separate task required the speeded identification of abstract iconic tactile messages which also emphasized either spatial or nonspatial processing. The visual task stimuli and tactile icons were presented simultaneously and nearly completely overlapped, thus requiring concurrent processing of the encoded information in each display. Performance on each of the two tasks (tactile icon identification and lane navigation) was measured in isolation in control blocks so that the decrements in performance in dual-task conditions could be directly compared. This comparison should provide insight into the role of code in visual/tactile processing interference.

It was expected that performance on the visual and tactile icon tasks would suffer a decrement when conducted concurrently, when compared to performance in isolation. The size of the decrement was expected to be larger when the tasks required the same processing code (either both spatial or both nonspatial processing). Subjective ratings of the difficulty in conducting each dual-task pairing were expected to correlate with objective performance patterns. Finally, due to the additional spatial processing requirements inherent to piloting the vehicle in these studies, performance decrements were expected to be greater when both tasks required spatial processing, compared to those when both required nonspatial processing.

METHOD

In the following sections, two experiments will be described that differ with respect to the encoding method (spatial versus nonspatial) for the tactile icons in an icon identification task. The participant populations for each experiment consisted of twenty-five students from the University of Michigan (sixteen males and nine females in each study, ages 19-28). Five (two males, three females) participated in both studies. All participants had normal or corrected-to-normal vision, no known injuries or disorders that may have affected tactile sensitivity, and a valid US driver's license. Participants drove a simulated vehicle in a scenario created in STISIM Drive™, a medium-fidelity desktop driving simulator, with a force-feedback steering wheel and floor-mounted throttle and brake pedals. Four circular (8 mm diameter) buttons embedded in the steering wheel face (two on each side, 40 mm apart) and two paddle triggers on the back side of the steering wheel were easily reached while gripping the wheel along its horizontal diameter and were employed for responses to the tactile icon identification task. An adjustable elastic support wrap was fastened around the torso, securing a three-by-three array of vibrotactile devices (hereon referred to as "tactors") to the mid-to-lower back, with the center column of tactors along the spine. The thickness of this wrap, and engine sounds played from a high-powered subwoofer assured that the auditory sound of tactor activation was masked. The tactors were pancake pager motors (model# 12820MD; M.P. Jones Associates, http://www.mpja.com) which were 14 mm in diameter, operated at frequencies of 80 Hz, and were driven with a wireless control unit developed by Jones and colleagues (Jones et al., 2006). The horizontal and vertical separation between each tactor was 100 mm, and the icons which required spatial discrimination (the spatially-encoded icons in Experiment 1) involved only the four tactors at the corners of the array (200 mm apart), and the center tactor (141 mm from each corner). These distances are considerably larger than the two-point discrimination threshold of 11 mm for vibrotactile stimuli on the back (Eskildsen, Morris, Collins, & Bach-Y-Rita, 1969).

The procedure for the two experiments was identical. Participants were first introduced to the driving simulator and the tactile icons they would identify during the study (the type of icons differed between studies). Afterward, a 15-minute training session introduced participants to each of three tasks (tactile icon identification, visual-spatial, and visual-nonspatial), which they performed in the driving simulation. Participants completed at least ten trials for each task in isolation, and were required to respond correctly in five successive trials at the end of each training segment. The participants then completed exactly five trials each for two training segments that paired the tactile task with each of the two visual tasks.

Next, participants completed a 30-35 minute experimental session, which consisted of five different task conditions. First, participants completed three 6-minute single-task blocks (one for each task, presented in randomized order) in succession. After an optional short break, they completed two dual-task blocks which paired the tactile task with each of the two visual tasks (also in randomized order). Finally, participants completed a 10-minute debriefing questionnaire which helped experimenters gain insight into participants' strategies and perceived difficulties for each of the task conditions. Completion of the entire experiment took approximately one hour, for which participants were compensated $10.

Driving Scenario

The scenario for each task condition was set on a four-lane road with alternating open stretches and "obstacle zones" where experimental trials occurred (24 trials per task condition). These obstacle zones included longitudinal parallel barriers which prohibited changing lanes while inside the zone. In the visual single-task conditions and dual-task conditions, three of the four lanes were obstructed (see Figure 2.6), and participants were instructed to enter the single unobstructed lane (the location of which was randomized). Limited visibility required participants to interpret visual images that were presented as if on a head-up display (HUD) in order to determine which upcoming lane was unobstructed (see descriptions in the 'Visual Tasks' section). In the single-task condition that involved only the tactile icon identification task, all four obstacle zone lanes were left unobstructed. Tactile and visual task stimuli were presented immediately after passing obstructions and the entirety of the presentation occurred within the latter half of the obstacle zones where no driving actions were required beyond maintaining lane position. Experimental trials occurred approximately every 15 s and lasted for 2.25 s.

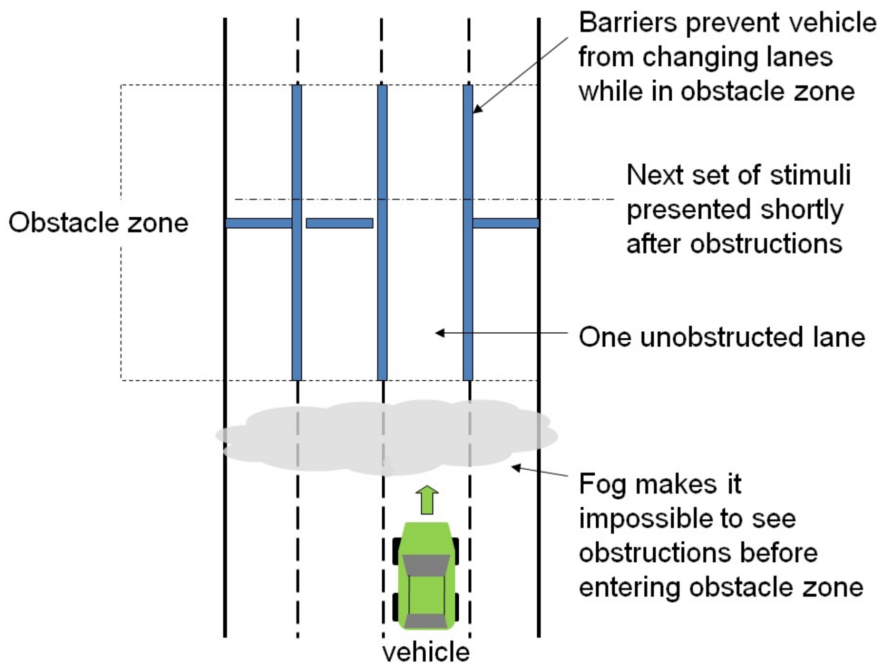


Figure 2.6: Schematic of an "obstacle zone" in the driving scenario.

Visual Tasks

The two visual tasks required participants to use information in three successive images to deduce the unobstructed lane for an upcoming obstacle zone. These tasks were designed to be similar in difficulty and differed mainly in the primary information processing code resources (spatial or nonspatial) they required. The performance measure for each visual task was accuracy of lane choice, i.e., the percentage of obstacle zones in which the participant entered the single unobstructed lane. When participants entered an incorrect lane, they were instructed to drive through the barrier, which temporarily stopped the vehicle and played a crash sound, then continue driving in the chosen lane while waiting for presentation of the next set of stimuli.

Each of the three visual task images was presented for 750 ms as if on a head-up display (see Figure 2.7). The visual-spatial task images were overhead views of the upcoming obstacle zone, each showing one of the four lanes as obstructed (in random order), leaving one to be identified as unobstructed. The images for the visual-nonspatial task were three color-coded rectangles (red, blue, green, or yellow), and participants were required to determine which of the four colors was not displayed. This color was later associated with the upcoming unobstructed lane when, after exiting the obstacle zone, a ”key” labeling the lanes with the four colors in a randomized order was displayed for two seconds. Each trial displayed a new key.

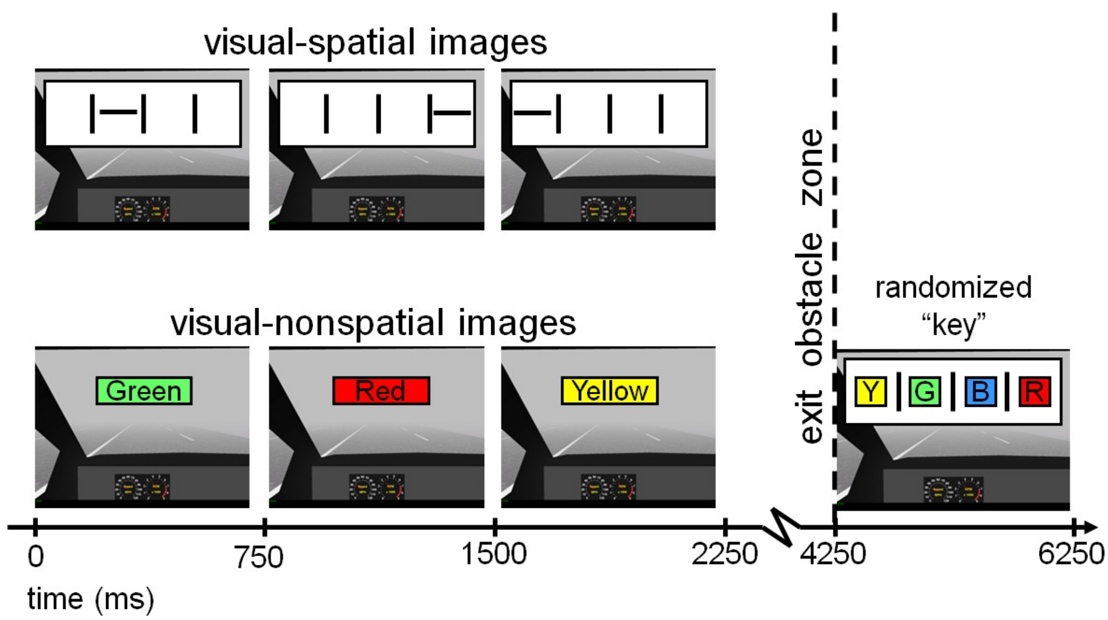


Figure 2.7: Example images displayed for each of the two visual tasks. Each image set indicates the third lane from the left as unobstructed in an upcoming obstacle zone.

After the three images were displayed, travelling the remaining length of the obstacle zone took at least two seconds (dependent on driving speed), during which time participants were unable to change lanes. In dual-task conditions, tactile icon presentation coincided with the image presentations, and this two-second window allowed ample time to respond to tactile icons while the steering wheel was at the home-position to preserve spatial mapping of the buttons. It also forced a delay in response to the visual task (changing lateral position of the vehicle), and since response time was measured for the tactile task but not the visual tasks, participants were encouraged (but not required) to respond to tactile icons first. Task interference, therefore, was designed to result from participants interpreting and responding to tactile icons while concurrently deducing and holding in working memory either a) the spatial information about which of four lanes is unobstructed or b) the nonspatial information about which of four colors was not shown.

Tactile Icon Identification Task

The tactile icons in the two experiments were described to participants as a display of information related to a hypothetical in-vehicle task, such as the logging of vehicle gas mileage. It was emphasized that the tactile task was unrelated to the task of piloting the vehicle, but that correctly responding to the tactile icons was equally as important as navigating the vehicle into the unobstructed lane specified by the visual task images. The icons used in each experiment were designed to be as similar as possible (e.g., equivalent interpretation difficulty, the same frequency, intensity, and total presentation duration), differing only in the encoding method – via spatial or nonspatial patterns – and thus the processing code required for their interpretation.

As in the visual task presentations, each tactile icon was composed of a sequence of vibration presentations over a period of 2250 ms total, which in dual-task conditions precisely overlapped the display of the visual task images. Participants were instructed to respond to each icon as soon as possible after it had been displayed in its entirety by pressing a two-button sequence on, or pulling a trigger behind, the steering wheel. Incorrect responses were announced with a system beep sound that was played at the beginning of the following obstacle zone, where there were no task-related presentations or required responses. The performance measures for this task were identification accuracy (percent correct buttons/trigger responses), and response time to the first button press or trigger pull. Because button-press responses were more complicated than trigger responses, times for the two response types were analyzed separately if found to be significantly different. Only response time data from correct responses were considered.

Spatially-Encoded Tactile Icons

The tactile icons in Experiment 1 were defined by the spatial relationship between three sequenced vibrations from individual tactors of the array attached to the back. The first and third vibrations of the sequence were presented via tactors at one of the four corner locations. The second was always presented from the center tactor, to serve as a consistent spatial reference point. Each of the three vibrations lasted 600 ms, with an interstimulus interval (ISI) of 150 ms. With short ISIs, vibrotactile masking is always a concern, but with comparable ISIs between a signal and a subsequent masking stimulus, performance in extracting a single dimension of information has been shown not to differ significantly from that in absence of a masker (Evans, 1987).

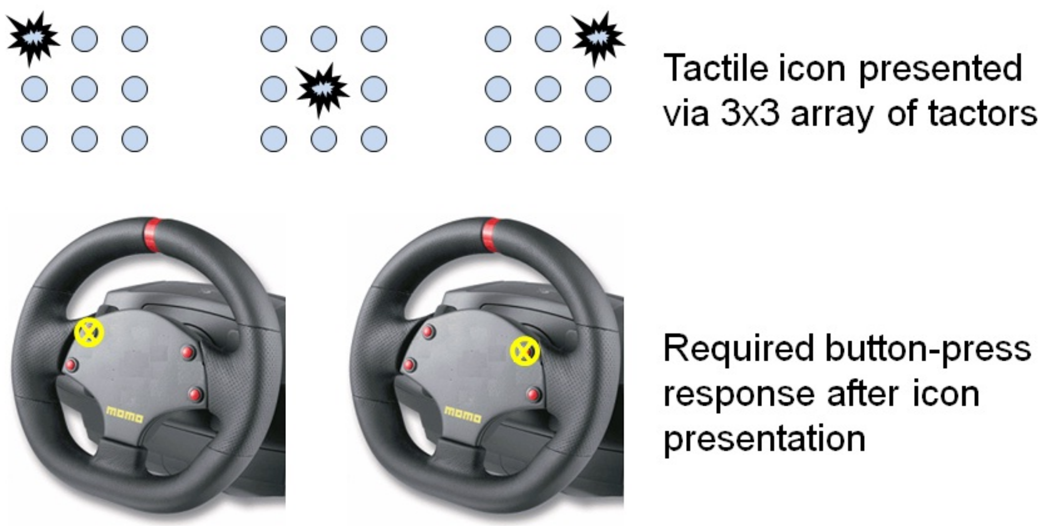


Figure 2.8: Example of the vibration sequence for a spatial tactile icon, and the required button-press response.

Participants were instructed to respond to the tactile icon as soon as possible following the onset of the third vibration. This represents the earliest point that all required information had been presented for icon identification, and response time was measured from this point. If the first and third vibrations originated from different corners of the array, participants were instructed to press, in order, the buttons on the steering wheel which mapped to the corresponding corners, as shown in Figure 2.8. If the same location was vibrated for the first and third vibrations, participants were instructed instead to pull the trigger mounted behind the wheel on the same side of space (left or right, regardless of whether the common location was in the top or bottom row). Trigger responses were required in one-third of tactile icon presentations and served to enforce that participants receive and interpret the entire icon before initiating a response. This assured a consistent zero point for measuring response time and also forced that interpretation of the icon involved a mental operation to decode the "message", thus engaging spatial processing resources.

Nonspatially-Encoded Tactile Icons

The tactile icons in Experiment 2 were defined by the numbers of pulses presented in two vibration "segments" (one, two, three, or four pulses per segment). All vibrations were presented from both outside tactors of the middle row of the 3x3 array. Figure 2.9 shows how the pulse on-times were divided evenly over each 900 ms segment, with 100 ms off-times between pulses. Pilot testing showed all four patterns to be easily and reliably distinguishable. The two segments were separated by 450 ms of off-time, so the entire icon presentation took 2250 ms.

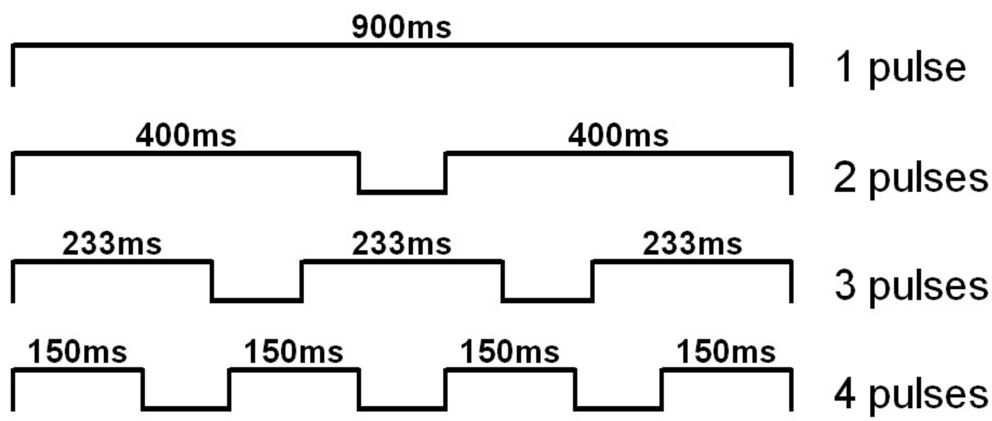


Figure 2.9: Four pulse patterns which could be displayed for a vibration segment of a nonspatial tactile icon. The off-time between pulses was uniformly 100 ms.

For this study, the four buttons embedded in the face of the steering wheel (Figure 2.8) were labeled with the numbers 1 and 2 on the top left and right, and 3 and 4 on the bottom left and right. Participants were instructed that if the two vibration segments for each icon were composed of different numbers of pulses, they were to press the buttons which corresponded to each pulse count, in order. If the two segments were composed of the same number of pulses, participants were instructed to instead pull the trigger behind the wheel on the same side as the number label (left trigger for odd numbers of pulses, right for even numbers). As in Experiment 1, trigger responses were required in 1/3 of the trials and enforced that the entirety of the icon's pattern was considered so that a comparison operation (with the nonspatial information "number of pulses" being compared) was made before participants initiated a response.

In contrast to the spatially-encoded icons in Experiment 1, identification of the nonspatial icons required participants to wait until nearly the entirety of the second segment had been presented to have all the necessary information to make the correct response. Although earlier identification was theoretically possible by judging the pulse duration instead of counting pulses, questions about task strategy on the debriefing questionnaire showed that participants largely used the prescribed strategy of counting pulses, and responded to the icon after completion of the second segment. Therefore, response time was measured from the end of the second vibration segment to the first button press or trigger pull.

Data Removal Procedures

In the debriefing for both experiments, participants were asked about strategies they employed when conducting each task. This was important because adopting certain strategies could serve to transform a task from one that consumes primarily one type of processing resource to one that stresses the other. For example, some participants described assigning numbers to the four lanes during the visual-spatial task, then instead of deducing the open lane by constructing the mental image of the relative spatial locations of the blocked lanes, they would mentally rehearse the numbers of the blocked lanes and determine which number was missing. Since this would allow completion of the task without fully engaging the intended processing code (and would not result in the intended competition for processing code resources), datasets from four participants who adopted these types of strategies were removed prior to analysis.

Despite strict training requirements of 10 trials and at least 5 consecutive correct responses for each single-task condition, two participants made large numbers of errors during one single-task experimental block due to initial confusions about required responses. In each case, performance in this single-task condition was more than 3 standard deviations below the mean, while performance in all other conditions was much closer to the mean. The result was a dramatic and artificial skewing of measures of relative dual-task performance; therefore datasets from these participants were removed. Following the removal procedures, additional participants were recruited so that for each experiment, 25 complete data sets were available for analysis.

RESULTS

Two primary methods were used to analyze the data from each experiment. When the assumptions of population distribution normality and variance homogeneity were validated, repeated measures linear models (using the General Linear Model formulation in SPSS 16.0) were used to identify main effects for each performance measure, and two-tailed Fisher's LSD post-hoc tests were used to determine differences between means. When data were not normally distributed (as was the case with visual task data in single-task conditions, due to ceiling effects), nonparametric Friedman tests were used to determine the overall effect of task condition. For simplicity, the five task conditions will be further abbreviated in the text with the notations listed in Table 2.1.

Notation	**Task Condition**
T	Tactile icon single-task
S	Visual-spatial single-task
N	Visual-nonspatial single-task
TS	Tactile icon/visual-spatial dual-task
TN	Tactile icon/visual-nonspatial dual-task

Table 2.1: Summary of task condition notations.

Experiment 1: Spatially-Encoded Icons

Tactile Icon Identification Performance. Figure 2.10a shows accuracy, and Figure 2.10b shows response times, for tactile icon identification for the *T* single-task condition and the *TS* and *TN* dual-task conditions. No significant difference was found for either accuracy or response time between icons which required a button-press and those which required a trigger response, therefore results from each response type were combined for analysis.

Task condition was found to significantly affect icon identification accuracy ($F(2,23) = 9.411$; $p = .001$). Post-hoc tests showed that accuracy in the *TS* condition (85.7%) was significantly lower than in the single-task *T* condition (93.3%; $p = .001$) and the *TN* condition (93.5%; $p < .001$). Accuracies in the *T* and *TN* conditions did not differ.

Task condition also had a significant effect on response times to correctly-identified icons ($F(2,23) = 27.004$; $p < .001$). Pairwise comparisons showed significant differences between all three conditions. The *TN* condition showed slightly longer response times (mean RT: 1189 ms) than the *T* condition (1112 ms; $p = .013$), and RTs in the *TS* condition (1340 ms) were considerably longer than in each other condition ($p <.001$ for both comparisons).

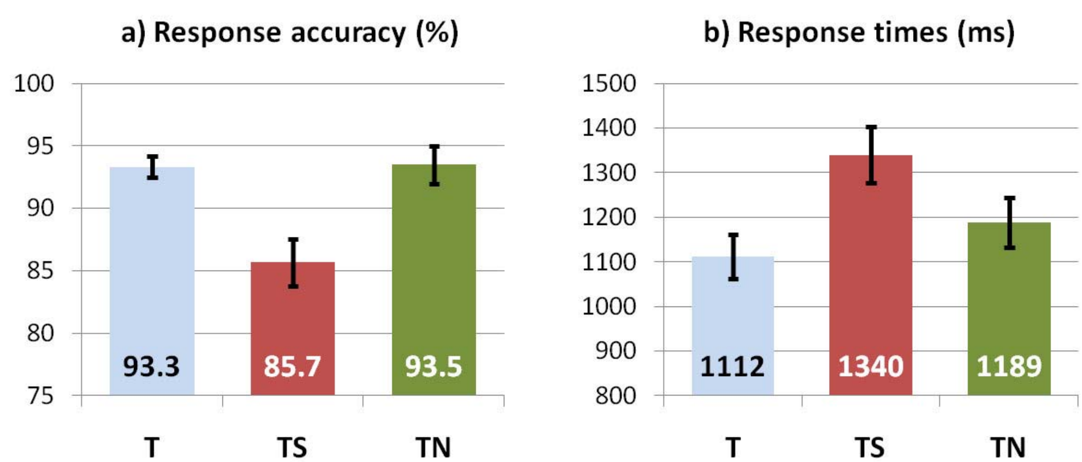


Figure 2.10: Response accuracies (a) and times (b) for the identification of spatially-encoded tactile icons in the single-task T condition and the dual-task TS and TN conditions in Experiment 1. Error bars represent standard error. Performance was worst in the TS condition for each measure.

Visual Task Performance. Figure 2.11 shows the percentage of correct lane choices based on accurate interpretation of visual task images. Lane choice accuracy was extremely high in single-task conditions (19 of the 25 participants chose every lane correctly in both single-task conditions, only 1 participant made as many as two incorrect choices), and the data were not normally distributed. Friedman tests showed a significant main effect of task condition on lane choice accuracy ($\chi2(3,N=25) = 59.322$; $p < .001$). Friedman tests also showed that dual-task accuracy was significantly worse than single-task accuracy for both the visual-spatial task ($\chi2(1,N=25) = 24.000$; $p < .001$) and visual-nonspatial task ($\chi2(1,N=25) = 17.000$; $p < .001$). Accuracies in the single-task conditions S (99.5%) and N (99.3%) did not differ significantly. Accuracies in dual-task conditions were normally distributed, and a repeated measures ANOVA found that they were significantly lower in the TS (79.7%) condition than in the TN (91.2%) condition ($F(1,24) = 20.353$; $p < .001$).

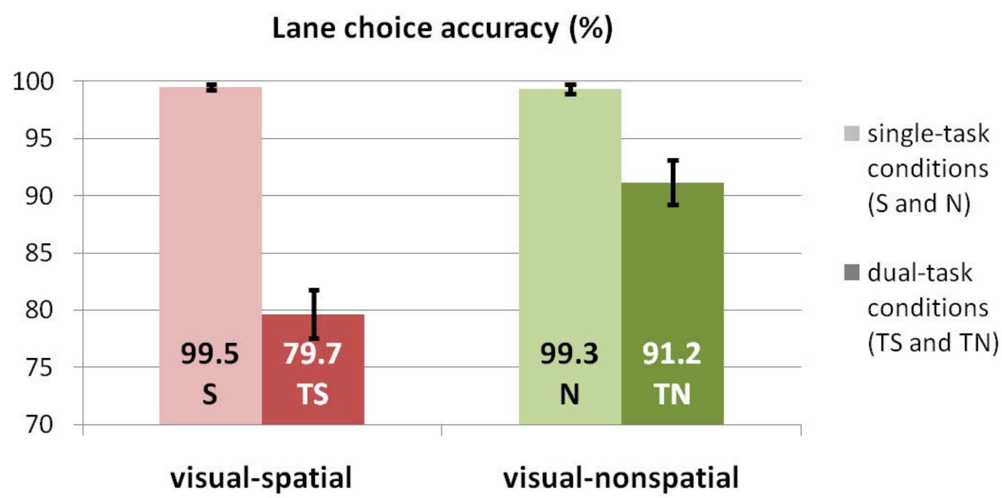


Figure 2.11: Visual task accuracies (% correct lane choice) for the visual-spatial and visual-nonspatial tasks in each respective single- and dual-task condition in Experiment 1 (spatially-encoded icons). Error bars represent standard error. Performance in each dual-task condition was significantly worse than in single-task conditions, with the worst overall performance in the TS condition.

Subjective Ratings. As part of the debriefing questionnaire, participants rated the difficulty of the two visual tasks when conducted in the *S* and *N* single-task conditions (to assure relatively equal difficulty) and when paired with the tactile icon identification task in the *TS* and *TN* dual-task conditions. Twenty-one of the twenty-five participants completed this part of the survey (others misinterpreted the questions or otherwise did not complete the ratings). Of these, 9 participants reported that the visual-spatial task was more difficult than the visual-nonspatial task in the single-task conditions, 7 reported that the visual-nonspatial task was more difficult, and 5 participants thought the two tasks were equally (not) difficult. In the dual-task conditions, however, almost all participants reported the visual-spatial task as more difficult (19 participants). Two participants who thought the visual-nonspatial task was more difficult as a single-task also felt it was harder in the dual-task condition. Participants were also asked to give a ratio of relative difficulty between the two visual tasks in the single- and dual-task conditions (e.g., "[X] was 1.3 times as difficult as [Y]"). The difficulty ratio between visual-spatial and visual-nonspatial increased from an average of 1.02 : 1 in the single-task conditions to 2.11 : 1 when paired with the tactile icon identification task in dual-task conditions.

Experiment 2: Nonspatially-Encoded Icons

Tactile Icon Identification Performance. Figures 2.12a and 2.12b show the accuracy and response times, respectively, for tactile icon identification for the *T* single-task condition and the *TS* and *TN* dual-task conditions. No significant difference was found between button and trigger responses for response accuracy, so the accuracy results were combined for analysis. In contrast to the findings of Experiment 1, the overall effect of task condition on icon identification accuracy did not reach significance ($F(2,23) = 3.159$; $p = .061$); however, a trend was observed where the drop in accuracy between the *T* (94.2%) and *TN* (88.7%) conditions was approximately twice as large (5.5%) as that of that between the *T* and *TS* (91.5%) conditions (2.7%).

Significant effects were found for response type ($F(1,24) = 5.968$; $p = .022$) and task condition ($F(2,23) = 17.124$; $p < .001$) on response time to correctly identified tactile icons. Trigger responses (mean RT: 1119 ms) were found to be slightly but significantly faster than button-press responses (1184 ms), therefore post-hoc analyses regarding the effect of task condition were separated by response type. For button-press responses, response times in the single-task *T* condition (1074 ms) were significantly faster than in the *TS* (1230 ms; $p = .001$) and *TN* (1247 ms; $p = .003$) conditions. Trigger responses showed the same differences, with response times in the *T* condition (1009 ms) significantly faster than in both the *TS* (1165 ms; $p < .001$) and *TN* (1183 ms; $p = .006$) conditions. Response times to the two dual-task conditions did not differ for either response type.

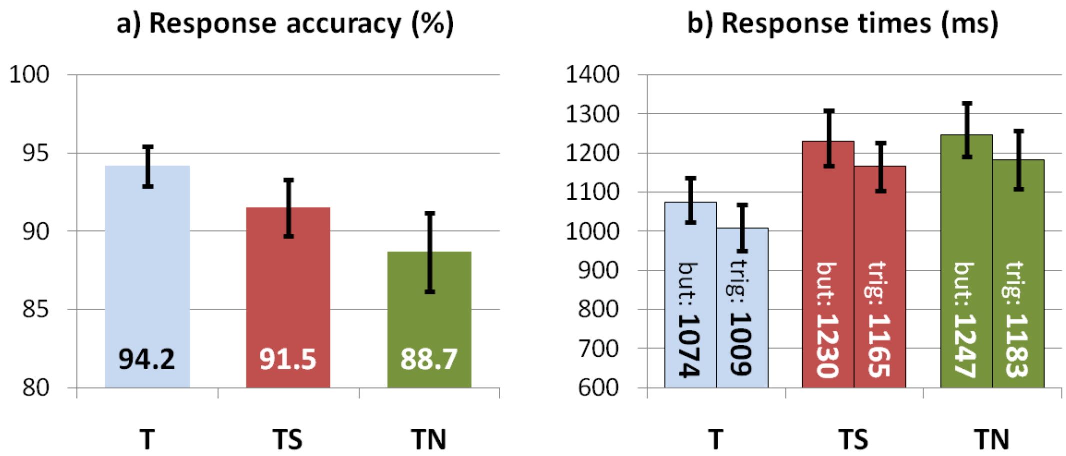


Figure 2.12: Response accuracies (a) and times (b) for the identification of nonspatially-encoded tactile icons in the single-task *T* condition and the dual-task *TS* and *TN* conditions in Experiment 2. Error bars represent standard error. Though the main effect did not reach significance, a trend shows the worst accuracy in the *TN* condition. Both dual-task conditions showed significantly longer response times than the *T* condition.

Visual Task Performance. As in Experiment 1, lane choice accuracy data were not normally distributed for the single-task conditions. A Friedman test showed a significant main effect of task condition ($\chi^2(3,N=25) = 55.091$; $p < .001$). Again, accuracy was lower in dual-task conditions than single-task conditions (visual-spatial: $\chi^2(1,N=25) = 19.000$; $p < .001$; visual-nonspatial: $\chi^2(1,N=25) = 23.000$; $p < .001$). Also mirroring the results in Experiment 1, accuracies in single-task conditions *S* (99.8%) and *N* (99.2%) did not differ (see Figure 2.13). The dual-task condition data, which were found to be normally distributed, were also found to be significantly different in a repeated measures model, though in the direction opposite that found in Experiment 1: accuracy in the *TN* condition (84.7%) was significantly lower than in the *TS* condition (93.2%; $F(1,24) = 19.067$; $p < .001$).

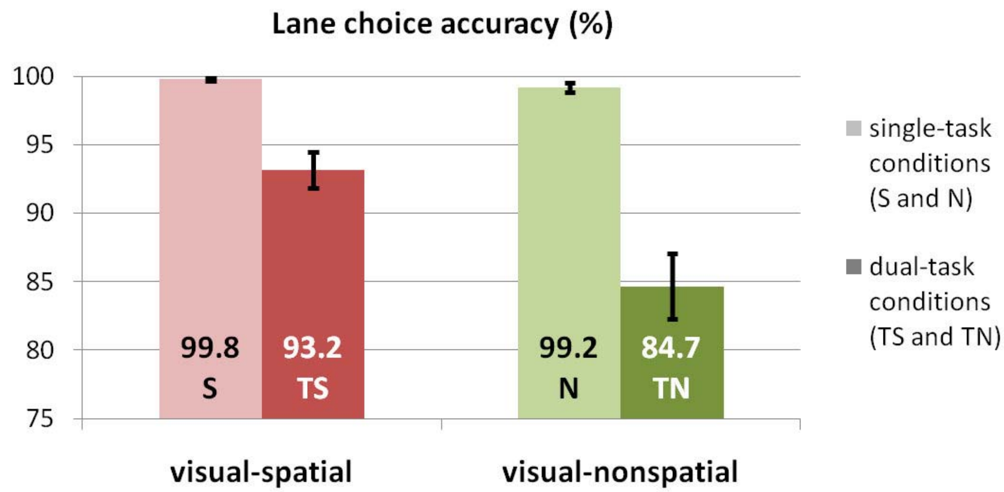


Figure 2.13: Visual task accuracies (% correct lane choice) for the visual-spatial and visual-nonspatial task in each respective single- and dual-task condition in Experiment 2 (nonspatially-encoded icons). Error bars represent standard error. Performance in each dual-task condition was significantly worse than in single-task conditions, with the worst overall performance in the *TN* condition.

Subjective Ratings. Twenty-four participants provided proper visual task difficulty ratings in the debriefing questionnaire (one participant apparently misinterpreted the questions). In contrast to the participant ratings from Experiment 1, participants generally thought the visual-nonspatial task was more difficult, with 13 participants reporting it to be more difficult than the visual-spatial task in the single-task conditions. 6 participants felt the visual-spatial task was more difficult, and 5 reported both tasks as equivalently difficult in isolation. When paired with the tactile icon identification task, 15 reported the visual-nonspatial task was more difficult, 7 reported the visual-spatial as more difficult, and 2 felt they were equally difficult. The average difficulty ratio between visual-spatial and visual-nonspatial single-task conditions was 1 : 1.16. In dual-task conditions, the visual-nonspatial task increased in relative difficulty, with an average difficulty ratio of 1 : 1.38.

Comparing the Results of Experiments 1 and 2: Dual-Task Performance Metrics

One primary goal in conducting the two reported experiments was to be able to quantify the effect of processing code interference in multitask performance with a given task set. While the results demonstrate how this interference may affect performance in each individual dependent measure, it is difficult to visualize the overall effect across all measures. Further, performance differences within individual measures may be muddied due to participants making tradeoffs between tasks in dual-task conditions. For example, despite instructions to participants to give equivalent effort to both tasks in dual-task conditions, debriefing discussions and some datasets made it evident that some participants tended to favor one task or the other in dual-task conditions. There was no consistent pattern as to which task was favored.

To account for these tradeoffs and additionally control for any speed/accuracy tradeoffs which may have been present within the tactile icon identification task, a metric was created that took into account performance across all dependent measures. This metric, p_{tV} (see Equation 2.1), calculated a participant's relative dual-task performance involving icon encoding method *t* and visual task *V* by dividing their performance levels for each dependent measure in dual-task conditions by their respective single-task levels. Relative performance in each dependent measure was then combined in a weighted sum, with weight coefficients determined by participant instructions (e.g., equivalent weights for the tactile and visual tasks, which were described to participants as equally important in dual-task conditions), and presentation frequency (e.g., coefficients of 2/3 and 1/3 for button and trigger RT measures, respectively). Higher *p* values represent better relative performance, and a value of 1.0 indicates equivalent performance in single- and dual-task conditions.

The metrics were calculated for each participant in each experiment, then compared in a repeated-measures ANOVA with tactile icon encoding method as a between-subjects variable. Table 2.2 shows the mean performance metrics for each combination of tactile encoding method and visual task. Analysis of the metrics showed a significant effect of visual task – worse dual-task performance with the visual-spatial (mean p = .886) than with the visual-nonspatial task (mean p = .914) ($F(1,48) = 8.103$; $p = .006$) –, and a strong interaction effect between icon encoding method and visual task ($F(1,48) = 57.008$; $p < .001$). Pairwise comparisons showed that the value of each performance metric in Table 2.2 was significantly different from the metric above/below it and to its left/right ($p < .005$ in all comparisons). Finally, the difference between performance metrics was calculated for each participant to estimate the magnitude of the performance decrement which could be attributed to processing code interference (rightmost column of Table 2.2). For the spatially-encoded icons, performance was 10.5% worse when both tasks required spatial processing resources. For nonspatially-encoded icons, performance was only 4.7% worse when both tasks required nonspatial processing resources. The difference between the decrements attributed to processing code interference was found to be significant in a pairwise t-test (t(24) = 2.820; p = .009).

$$p_{tV} = \left(\frac{1}{2} \left(\frac{1}{2} \left(\frac{2}{3} \left(\frac{RT_{but}(T_t)}{RT_{but}(T_tV)} \right) + \frac{1}{3} \left(\frac{RT_{trig}(T_t)}{RT_{trig}(T_tV)} \right) \right) + \frac{1}{2} \left(\frac{A_{tac}(T_tV)}{A_{tac}(T_t)} \right) \right) + \frac{1}{2} \left(\frac{A_{vis}(T_tV)}{A_{vis}(V)} \right) \right)$$

Equation 2.1: Metric for dual-task performance in conditions which paired visual-task *V* (<u>S</u>patial or <u>N</u>onspatial) and tactile icons defined by encoding method *t* (<u>s</u>patial or <u>n</u>onspatial). *RT*(X) = mean response time in task condition X; *but*, *trig* = button press, trigger responses; A_{tac}(X) = tactile icon accuracy; A_{vis}(X) = visual task accuracy.

Tactile icon encoding method	**Visual task**		**Decrement attributed to processing code interference**
	Visual-spatial	**Visual-nonspatial**	
Spatial	p_{sS} = .841 (15.9% decrement)	p_{sN} = .946 (5.4% decrement)	10.5%
Nonspatial	p_{nS} = .930 (7.0% decrement)	p_{nN} = .883 (11.7% decrement)	4.7%

Table 2.2: Dual-task performance metrics for each combination of tactile icon and visual task.

DISCUSSION

Recent years have seen considerable interest in communication via the sense of touch due, largely, to an increasing risk of visual and auditory data overload in many complex domains, such as aviation, medicine, or the car cockpit. Multiple Resource Theory (MRT) is one model commonly cited in support of the use of tactile displays for such environments, asserting that presenting task-relevant stimuli via separate sensory channels can reduce processing interference and lead to improved overall multitask performance. However, the effectiveness of this approach for supporting timesharing likely depends on additional factors, such as the extent to which tasks involve the same processing stage and share information processing codes. The studies described in this manuscript focus on the latter aspect and investigate the extent to which processing code interference affects performance in concurrent visual and tactile tasks. The reported findings quantify the effect of processing code interference, which is a critical step toward building computational models of human cognition and performance that can guide the design of tactile displays. These results also shed light on the challenges that later-stage cognitive bottlenecks – those that occur "downstream" of tactile perception – pose to tactile information processing (Gallace et al., 2007a; MacLean, 2008b).

Performance in interpreting tactile icons was shown to suffer – independent of whether encoded via spatial or nonspatial patterns – when concurrently conducting simple visual tasks that required either spatial or nonspatial cognitive processing resources. However, the negative effect on performance was significantly greater when spatially-encoded icons were paired with the visual-spatial task, and when nonspatially-encoded icons were paired with the visual-nonspatial task. Subjective ratings of task condition difficulty showed a similar pattern, with dual-task conditions that combined tasks engaging the same processing code being rated as most difficult. These findings confirm our expectations and support the MRT claim that processing interference is greater when tasks require the same processing code.

There was a notable difference in the size of the performance decrement between tasks that competed for spatial processing resources and those that competed for nonspatial resources. According to the dual-task performance metrics, performance when concurrently interpreting spatially-encoded icons and visual-spatial task images suffered 10.5% more than when interpreting the same icons with the visual-nonspatial task images. In contrast, the dual-task performance decrement with the nonspatial icons/visual-nonspatial task pairing was 4.7% greater than the decrement when the same icons were paired with the visual-spatial task. The roughly 6% difference between the spatial-spatial and nonspatial-nonspatial performance decrements, while statistically significant, may not matter very much operationally. However, the spatial processing code decrement was also more than double that of its nonspatial counterpart, and it is likely that such a difference could scale up in more complex real-world environments, thus making it much more relevant to design.

A number of factors may have contributed to the significantly larger difference for the spatial-spatial task pairing, and these factors are relevant to the discussion of introducing tactile displays to the driving environment. First, to some extent, it likely reflects a higher demand for spatial processing due to the additional processing requirements associated with piloting the simulated vehicle. The real-world task of driving requires continual assessment of the location of the vehicle along the route to their destination, and constant monitoring for changes in lane position and potential obstacles, such as other vehicles and pedestrians. While the simulated driving activity required in the reported studies was greatly simplified, some amount of spatial processing was still required to maintain the vehicle between the parallel lane barriers of the obstacle zone during the presentation of, and response to tactile and visual task stimuli. This spatial processing requirement would likely be a larger factor in more realistic driving scenarios.

Another related issue is that controlling a vehicle requires manual interaction (e.g., turning the wheel, adjusting pressure on the throttle/brakes) which stresses spatial resources at the response stage of processing (Wickens, 2002; Wickens, 2008). The current studies required manual control of the vehicle, and additionally required manual (button/trigger) responses to the tactile icons. Despite the careful control exercised through scenario design to delay response to visual task images until after responses to tactile icons (to minimize interference at the response stage), the act of responding to the icons may have interfered with ongoing processing of visual-spatial task stimuli. For example, in dual task conditions the spatial processing required to generate and execute motor programs in responding to tactile icons could coincide with the later cognitive stages of the visual task, e.g., mentally comparing/combining the visual task images to deduce the solution (open lane or the missing color) or mentally rehearsing the solution. In dual-task conditions where spatial processing resources are already heavily stressed (i.e., the spatial-spatial task pairing), this could contribute to a larger performance decrement.

If responses to the tactile icons in these studies had been made verbally, rather than manually, this may have instead imposed a greater processing interference on nonspatial processing resources, and led to larger performance decrements in the nonspatial-nonspatial task pairing. This assumption is supported by Brooks (1968), who showed worse performance on a mental imagery spatial task when question probes related to the task required manual responses, and worse performance on a mental verbal task when questions required verbal responses, compared to the opposite response methods. This demonstrates that processing code interference can affect performance on concurrent tasks even when they occupy separate processing stages (e.g., cognition and response stages) and suggests that the response mechanisms for existing tasks in an environment should be considered as well when introducing tactile display technologies. Currently, most tasks in the driving environment require manual interaction, which suggests employing nonspatial encoding methods for introduced tactile displays. However this suggestion may change as in-vehicle technologies increasingly employ voice-activated controls.

One other factor that may have contributed to the greater spatial-spatial performance decrement was that there may have been greater similarity between the spatial tactile and visual tasks than between their nonspatial counterparts. While the spatial tasks both required judgment of relative spatial locations, the nonspatial tasks involved two different categories of information: numbers and colors. It is likely that greater task interference, and thus worse performance, would have resulted if both tasks required processing sets of the same category of information. For example, recent work has shown that people encounter surprising difficulty when tasked to concurrently count the number of displayed visual stimuli and the number of body locations presented with vibrations (Gallace, Tan, & Spence, 2007b). In the real-world driving environment, where task sets involve processing several different categories of information, it is relatively unlikely that separate tasks would require concurrently processing the same category of nonspatial information. This means that when nonspatial processing code interference *does* occur, it is likely to involve separate categories of information. Since the results of the current studies suggest this type of interference may lead to relatively smaller performance decrements, this also supports employing nonspatial encoding methods for tactile displays in the driving environment.

Finally, in the discussion of the processing requirements for existing tasks in the driving environment, it is important to note that it may be easier for drivers to avoid situations in which concurrent nonspatial processing is required. The types of in-vehicle tasks that commonly require nonspatial processing – e.g., reading street or traffic signs, engaging in conversations, or listening to music – can usually be abandoned or temporarily postponed if nonspatial processing resources are needed for more critical tasks. In contrast, the safety-critical task of piloting the vehicle – which requires primarily spatial processing – cannot be as easily interrupted. For these reasons, nonspatial tactile displays may naturally face less processing resource competition and would therefore be preferable.

If the minimization of processing interference was the only design consideration, it would be easy to recommend nonspatial over spatial encoding methods for complex tactile displays being introduced into the driving environment. However, it needs to be made clear that this should only be one of many factors that are considered in design. Another performance-relevant factor which was not investigated here is the interpretability of the tactile pattern, which depends on the amount of information conveyed tactually and the semantic mapping between signal modulation and encoded message. This factor is especially important given that the tactile modality does not lend itself well to the presentation of some information formats which are effectively communicated via vision/audition. For example, there is ample evidence that displaying navigation information visually or auditorily via verbal directions (e.g., "turn right onto Main Street") leads to better navigation performance and fewer driving errors than displaying a map of the desired route (Wetherell, 1979; Wickens & Hollands, 2000). If the tactile channel is employed for display of navigation information, however, tactile patterns relaying such complex verbal messages may be considerably more difficult to interpret than a spatial representation. Indeed, some notable designs for in-vehicle tactile navigation displays have demonstrated success with spatially-encoded signals, which naturally map to the spatial directions being presented (e.g., Kern, Marshall, Hornecker, Rogers, & Schmidt, 2009; van Erp & van Veen, 2004; van Erp, van Veen; Jansen, & Dobbins, 2005). Ultimately, the pursuit of "transparent" tactile communication – that which places minimal load on attentional resources – will depend on continued efforts to find the most natural/intuitive mappings between tactile patterns and communicated information (Rupert, 2000), as well as those that minimally compete with an existing task set for processing resources.

REFERENCES

Brewster, S. & Brown, L.M. Tactons: Structured tactile messages for non-visual information display. *Proc. 5th Australasian User Interface Conference (AUIC2004)*, 15-23, 2004.

Brewster, S. & King, A. An investigation into the use of tactons to present progress information. *Lecture Notes in Computer Science, 3585*, 6-17, 2005.

Brewster, S., Wall, S., Brown, L., & Hoggan, E. Tactile Displays. In: The Engineering Handbook on Smart Technology for Aging, Disability and Independence. Computer Engineering Series, John Wiley & Sons, Chichester, 2008.

Brooks, L.R. Spatial and verbal components of the act of recall. *Canad. J. Psychol./Rev. Canad. Psychol., 22*(5): 349-368, 1968.

Brown, L.M., Brewster, S.A. & Purchase, H.C. Multidimensional tactons for non-visual information display in mobile devices. *Proc MobileHCI'06*, 231-238, 2006.

Calhoun, G.L., Fontejon, J.V., Draper, M.H., Ruff, H.A., & Guilfoos, B.J. Tactile versus aural redundant alert cues for UAV control applications. *Proceedings of HFES 48th Annual Meeting*, 137-141, 2004.

Chan, A., MacLean, K.E., & McGrenere, J., Learning and identifying haptic icons under workload. *Proc. First World Haptics Conference (WHC '05)*, 432-439, 2005.

Craig, J.C. Vibrotactile pattern perception: Extraordinary observers. *Science*, *196*, 450-452, 1977.

Eskildsen, P., Morris, A., Collins, C.C., & Bach-Y-Rita, P. Simultaneous and successive cutaneous two-point threshold for vibration, *Psychonomic Science, 14*: 146–147, 1969.

Evans, P.M. Vibrotactile masking: Temporal integration, persistence, and strengths of representations. *Perception & Psychophysics, 42*(6): 515-525, 1987.

Ferris, T., Penfold, R., Hameed, S., & Sarter, N. The implications of crossmodal links in attention for the design of multimodal interfaces: A driving simulation study. *Proceedings of the HFES 50th Annual Meeting*, 1481-1485, 2007.

Ford, S., Daniels, J., Lim, J., Koval, V., Dumont, G., Schwarz, S.K.W., & Ansermino, J.M. A novel vibrotactile display to improve the performance of anesthesiologists in a simulated critical incident. *Anesthesia & Analgesia, 106*: 1182-1188, 2008.

Gallace, A., Tan, H.Z., & Spence, C. The body surface as a communication system: The state of the art after 50 years. *Presence, 16*(6): 655-676, 2007a.

Gallace, A., Tan, H.Z., & Spence, C. Multisensory numerosity judgments for visual and tactile stimuli. *Perception & Psychophysics, 69*(4): 487-501, 2007b.

Geldard, F.A. Some neglected possibilities of communication. *Science, 131*: 1583-1588, 1960.

Hameed, S., Ferris, T., Jayaraman, S. & Sarter, N. Using informative peripheral visual and tactile cues to support task and interruption management. *Human Factors, 51*(2): 126-135, 2009.

Ho, C., Tan, H.Z. & Spence, C. Using spatial vibrotactile cues to direct visual attention in driving scenes. *Transportation Research Part F, 8*: 397-412, 2005.

Hoggan, E. & Brewster, S. New parameters for tacton design. *Proc. SIGCHI Human Factors in Computing Systems*, 2417-2422, 2007.

Jones, L.A. & Sarter, N.B. Tactile displays: Guidance for their design and application. *Human Factors, 50*(1): 90-111, 2008.

Jones, L.A., Lockyer, B. & Piateski, E. Tactile display and vibrotactile pattern recognition on the torso. *Advanced Robotics, 20*(12): 1359-1374, 2006.

Kern, D., Marshall, P., Hornecker, E., Rogers, Y., & Schmidt, A. Enhancing navigation information with tactile output embedded in the steering wheel. *Lecture Notes in Computer Science, 5538*, 42-58, 2009.

MacLean, K.E. Haptic interaction design for everyday interfaces. *Reviews of Human Factors and Ergonomics*. Ed. M. Carswell. Santa Monica, CA: HFES, 2008a.

MacLean, K.E. Foundations of transparency in tactile information design. *IEEE Transactions on Haptics, 1*(2): 84-95, 2008b.

MacLean, K. & Enriquez, M. Perceptual design of haptic icons. *Proc Eurohaptics 2003,* 2003.

Pettitt, R.A., Redden, E.S., & Carstens, C.B. Comparison of army hand and arm signals to a covert tactile communication system in a dynamic environment. U.S. Army Research Laboratory Technical Report, 2006.

Rupert, A.H. An instrumentation solution for reducing spatial disorientation mishaps. *IEEE Engineering in Medicine and Biology Magazine, 19*(2): 71-80, 2000.

Sklar, A.E. & Sarter, N.B. Good vibrations: Tactile feedback in support of attention allocation and human-automation coordination in event-driven domains. *Human Factors, 41*(4): 543-552, 1999.

Tan, H.Z., Durlach, N.I., Reed, C.M., & Rabinowitz, W.M. Information transmission with a multifinger tactual display. *Perception & Psychophysics, 61*(6): 993-1008, 1999.

Tan, H.Z., Gray, R., Young, J.J. & Traylor, R. A haptic back display for attentional and directional cueing. *Haptics-e: The Electronic Journal of Haptics Research, 1*(3). 2003.

Ternes, D. & MacLean, K. Designing large sets of haptic icons with rhythm. *Proc Eurohaptics 2008*, 199-208, 2008.

van Erp, J.B.F. Guidelines for the use of vibro-tactile displays in human computer interaction. *Proc Eurohaptics 2002*, 18-22, 2002 van Erp, J.B.F., Groen, E.L., Bos, J.E. & van Veen, H.A.H.C., A tactile cockpit instrument supports the control of self-motion during spatial disorientation. *Human Factors, 48*(2): 219-228, 2006.

van Erp, J.B.F. & van Veen, H.A.H.C. Vibrotactile in-vehicle navigation system. *Transportation Research Part F, 7*, 247-256, 2004.

van Erp, J.B.F., van Veen, H.A.H.C., Jansen, C. & Dobbins, T. Waypoint navigation with a vibrotactile waist belt. *ACM Transactions on Applied Perception, 2*(2): 106-117, 2005.

Verillo, R.T., Fraioli, A.J. & Smith, R.L. Sensation magnitude of vibrotactile stimuli. *Perception & Psychophysics, 6*, 366-372, 1969.

Welch, R.B. & Warren, D.H. Intersensory interactions. In Boff, K., Kaufman, L., & Thomas, J. (Eds.) Handbook of Perception and Human Performance, Volume 1: Sensory Processes and Perception. John Wiley & Sons, Inc., 1986.

Wetherell, A. Short-term memory for verbal and graphic route information. *Proceedings of HFES 23rd Annual Meeting*, 464-468, 1979.

Wickens, C.D. Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*(2): 159-177, 2002.

Wickens, C.D. & Hollands, J.G. Engineering Psychology and Human Performance (3rd Ed). Prentice Hall, NJ, 2000.

Wickens, C.D. Multiple resources and mental workload. *Human Factors, 50*(3): 449-455, 2008.

Contributions to Tactification Design

Some of the findings described in the tactile change blindness study (Part A) and the processing code studies (Part B) were directly applicable to the tactification designs described in Chapter 5. First, because vibration intensity was determined to be one of the most naturally-mapping dimensions for encoding physiological data (see the next chapter), the tactifications employed this dimension and thus were susceptible to change blindness effects. Because the Gradual display condition showed such low change detection rates, the tactification designs included discrete steps in intensity, rather than analog intensity changes. Also, since the changes which were larger in magnitude (6 dB) were considerably more resilient than the smaller changes (3 dB) to change blindness effects, the larger changes were employed in the tactification design. Finally, since the tactifications were designed for and evaluated in a multitask scenario, it was of particular interest that change detection under the Mudsplash condition was significantly impacted by the presence of a secondary task. The complexity of the tactification display is most likely to reproduce conditions similar to the mudsplash, in that sequences of vibrations at one location of the body may coincide with important (intensity) changes presented to a different location. To ensure that any resulting effects of change blindness minimally impacted overall performance in detecting these changes, it was important to not rely solely on the detection of intensity changes in the tactification design. Therefore, the tactification designs employed redundant signal modulations – changes in both intensity and another dimension of the signal, such as the location or duration of presentation – to improve the likelihood of change detection.

Part B of this chapter showed how multitasking performance can suffer when tactile information processing and the processing required for concurrent tasks share the same processing code. Because the task set of an anesthesiologist is very complex and, at times, requires either or both the spatial and symbolic processing codes, the best way to that interference at the cognitive processing stage could be minimized was again to use redundancy in tactile encoding methods. Critical information that was included in the tactification displays was always encoded via a spatial encoding method and at least one nonspatial/symbolic method. By allowing anesthesiologists the option of interpreting the information by engaging either processing code (ideally, the one that is engaged to a lesser extent by concurrent tasks), interference at the cognitive stage could be minimized.

We have now examined two potential challenges to tactile information processing, interferences at the perceptual and cognitive stages of processing. By taking steps to minimize these interferences, we can satisfy the first property of preattentive reference displays: minimization of interference between display components and with concurrent tasks. The next chapters relate to the second property: communicating partial information on the state of the displayed system (the physiological health state of the patient) to support decision making about whether, and when, to shift attention between tasks. Chapter 3 details a survey of anesthesia providers which addressed this property by determining what types of information must be known, in regard to the state and dynamics of key physiological parameters, in order to make efficient task-switching decisions.

Chapter 3

Survey of Anesthesia Providers

This chapter summarizes the responses to an online survey of an international community of anesthesia providers (anesthesiologists and Certified Registered Nurse Anesthetists (CRNAs)). Goals of the survey included gathering input for prototype designs of vibrotactile physiological displays as well as for the experimental designs of the studies described in Chapters 5 and 6. The survey also served to ensure that these studies adequately addressed some of the real-world challenges faced by anesthesia providers in the task of physiological monitoring. In particular, contributions of this survey include:

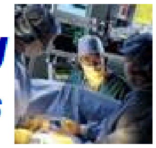
1) Verification of the need for more effective ways to display information related to invasive blood pressure measures, and the identification of perioperative patient care tasks and circumstances that pose the greatest challenges to effective monitoring of blood pressure and other physiological measures.
2) Identification of specific types of information which, when effectively displayed, are/would be most useful in supporting blood pressure management under these challenging conditions.
3) Information regarding attitudes about the display of blood pressure information via vibrotactile displays, and constraints and preferences to consider in the design of such displays.
4) Identification of natural mappings between types of blood pressure information and vibrotactile encoding mechanisms for relating the information.

5) Input regarding physiological measures, other than blood pressure, that could benefit from improvements in display methods.

Faculty and residents of the University of Michigan Department of Anesthesiology and members of the International Society for Technology in Anesthesia (STA) were sent invitations to participate in the study via moderated email lists. The survey consisted of 27 questions, and took roughly 20 minutes to complete. Of the roughly 500 anesthesia providers who received this invitation, 34 anonymous responses (6.8%) were collected. Of the respondents, 24 were anesthesiology faculty members (17 with 10 or more years of experience), 4 were anesthesiology residents, 5 were CRNAs, and 1 was a Human Factors researcher with expertise in the anesthesiology domain.

Figure 3.1 is an example screenshot from the online survey. As the figure shows, some questions supported the collection of quantitative data, such as ratings for interference levels that some tasks imposed. Others supported a “free response” format. In this chapter, the results which were most relevant to the research activities described in Chapters 5 and 6 are detailed. A complete list of all survey questions can be found in Appendix 1.

Supporting patient monitoring through novel tactile displays



4 **

Do any of the following routine tasks interfere with your ability to closely monitor critical patient parameters such as blood pressure?

Please rate each task listed below. You will have a chance to elaborate later.

Task	1 No interference	2 Slight interference	3 High interference
Setting up infusion pumps or replacing saline drip	○	◉	○
Documentation/entering records on Centricity (such as drugs administered)	○	◉	○
Administering a drug via syringe	◉	○	○
Draining Foley catheter bag	○	◉	○
Checking/replacing sensors on patient (such as pulse ox)	◉	○	○
Drawing blood from vein or A-line	○	◉	○
Intubation	○	○	◉
Train-of-4	◉	○	○
Preparing drug syringes for the current or a later case	○	◉	○

5 (optional)

If you can think of any other tasks that interfere with monitoring, please list them below.

```
placing lines for drug infusions.
```

Next Page

survey progress

Figure 3.1: Screenshot of the online survey. A full list of survey questions can be found in Appendix 1.

SURVEY RESPONSE SUMMARY

Survey responses were compiled and sorted into 5 categories, as listed on the first page of this chapter (page 73). Results are summarized below, as they relate to each of these categories.

1) Verification of the need for more effective ways to display information related to invasive blood pressure measures, and the identification of perioperative patient care tasks and circumstances that pose the greatest challenges to monitoring blood pressure and other physiological measures.

Responders expressed a general difficulty with monitoring blood pressure dynamics and patterns during the induction and emergence phases of anesthesia when they are required to multi-task. Several responders noted that blood pressure is especially critical to monitor closely during these phases, because of unpredictable responses to anesthetic drugs and the potential for overdosing a patient. Common co-morbidities such as hypertension, hypotension, cardiac and cerebrovascular diseases, and cardiac arrhythmias, and rare but especially problematic co-morbidities such as pheochromocytoma (a neuroendocrine tumor that causes hypertension that is resistant to treatment) were mentioned as being linked to a tendency for blood pressure to be more unstable, thus increasing the importance of paying close attention to pressure levels. Several responders noted challenges imposed by the current format of the display of blood pressure, namely visual displays that can be difficult to sample while conducting tasks, and auditory alarms that are largely ineffective because of high rates of false alarms. Some noted the common practice of setting blood pressure alarm limits at extreme values to minimize false alarm rates.

One question asked specifically about the extent to which some induction tasks interfere with the ability to closely monitor blood pressure (and other parameters). "Interference index" values were calculated for each of these tasks by adding 3 points to the index for "high interference" responses, and 1 point for "slight interference" responses (0 points for "no interference" responses). Figure 3.2 shows the interference index values for each task. Endotracheal intubation was identified as the task that interferes the most with blood pressure monitoring, followed by drawing blood, programming infusion pumps, and checking/replacing sensors.

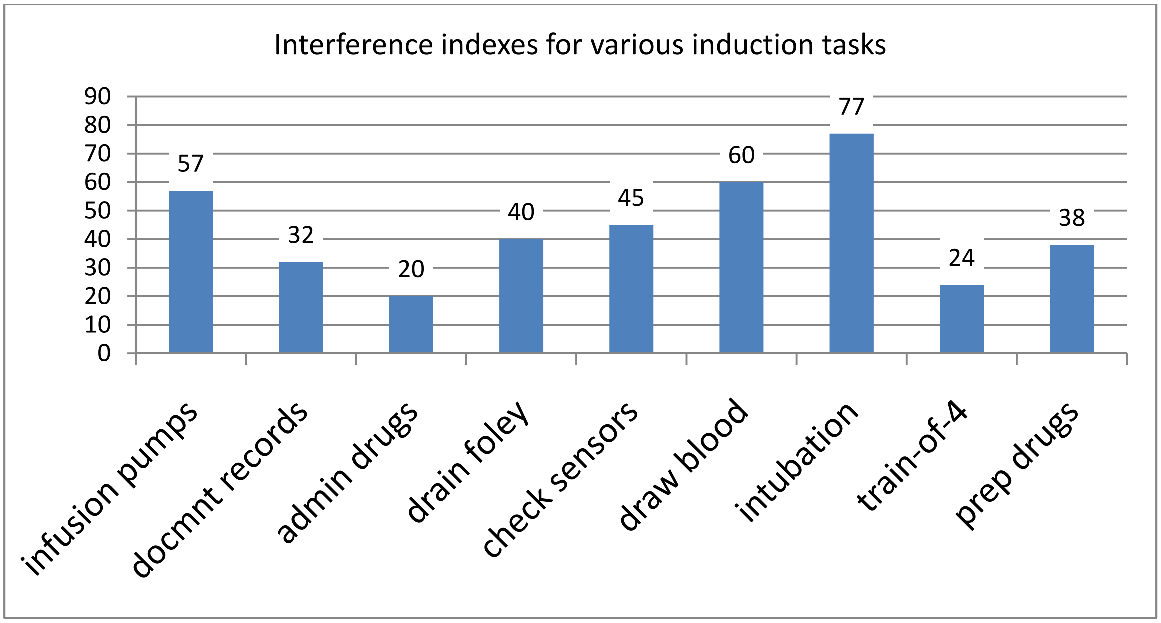


Figure 3.2: Responses to questions about how much conducting each of the above tasks interferes with the ability to monitor blood pressure.

Free responses about other tasks that routinely interfered with monitoring included: line placement, patient positioning procedures, maneuvering the OR table, placing the Bair Hugger (a body suit used for temperature regulation), and transesophageal echocardiography (TEE; ultrasound imaging of the heart that involves insertion of a transducer into the esophagus).

Free responses to questions about other circumstances that pose a challenge to effective monitoring frequently referenced auditory noise levels in the operating room (OR). Eight responses explicitly listed overall noise levels as being the primary source of interference with their ability to effectively monitor the patient. Sources of noise that were mentioned could be related to “necessary” activities, e.g., conversations between surgeons and other OR staff, noise from surgical tools, auditory monitors and notifications such as the pulse oximetry display and capnograph warnings and alerts. “Unnecessary” sources of noise were a common complaint, with 12 responses fitting this category, including inappropriate conversations between OR staff, loud and/or "offensive" music, and frequent false alarms from monitoring systems.

2) Identification of specific types of information which, when effectively displayed, are/would be most useful in supporting blood pressure management under these challenging conditions.

A set of questions asked survey participants to identify the three most useful types of information for supporting blood pressure management under routine and critical situations. Scores were tallied for each of these types of information by adding 3 points for each "most useful" response, 2 points for each "second-most useful" and 1 point for each "third-most useful" response. Write-in responses were allowed, and one participant noted that capnography measures, which weren't included on the original list, were extremely important. Figure 3.3 summarizes these "usefulness" scores under routine conditions. Responses for usefulness under critical incidents/situations (a free-response question) were not appreciably different, but in general expressed higher importance for access to each "useful" type of information.

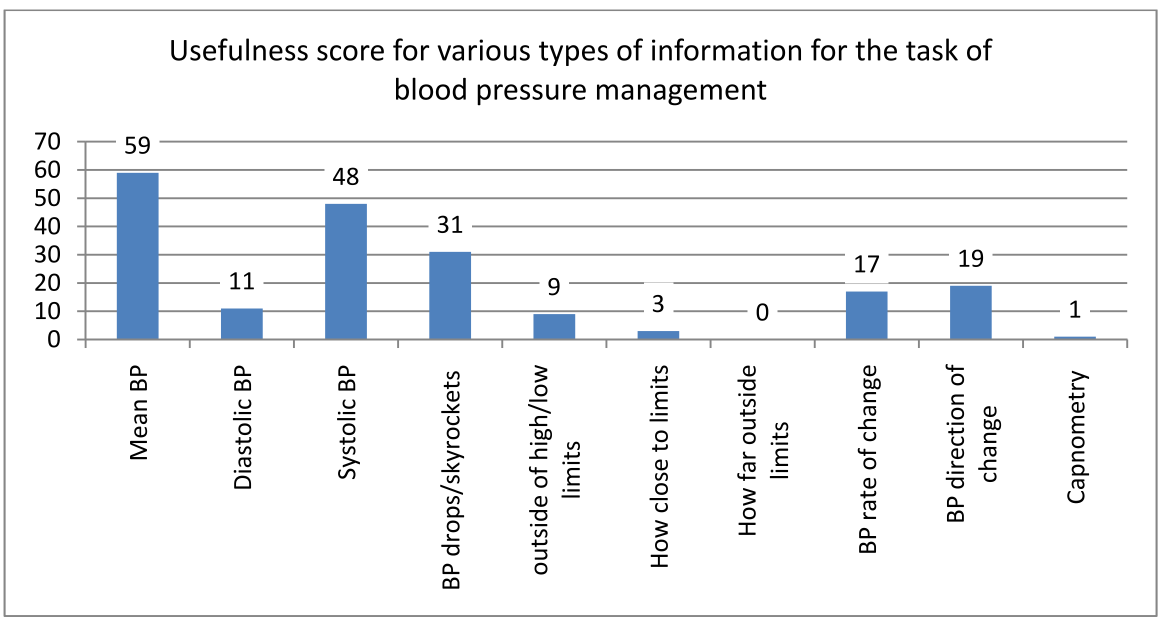


Figure 3.3: "Usefulness" scores for various types of information, relating how useful each is for the task of blood pressure management under routine conditions.

The top three most useful types of information – Mean Arterial Pressure (MAP), systolic blood pressure, and dramatic changes (drops/skyrockets) in blood pressure – are currently displayed in some format within most physiological monitoring systems (though in the latter case, which represents alarm conditions, the information is not displayed very effectively). Interestingly, the next two most useful types of information are not explicitly displayed in most monitoring systems: rate and direction of change in blood pressure. Under current display configurations, anesthesiologists need to keep mental track of changing blood pressure levels to infer these types of information. Alternatively, they can access a history chart of blood pressure, but the electronic charting system is frequently lagging behind by as much as several minutes.

3) Information regarding attitudes about the display of blood pressure information via vibrotactile displays, and constraints and preferences to consider in the design of such displays.

After introducing the idea of a vibrotactile display to redundantly display some of the more critical and/or useful types of blood pressure information, a series of questions asked about preferred implementations of this approach. Responders rated six body locations for the placement of a tactile display – wrist, forearm, upper arm, thigh, waist, and mid/lower back – in terms of three attributes: obstructiveness (how much harnesses around and/or vibrations presented to the body location would physically interfere with ongoing tasks), annoyance (the amount of frustration and/or physical discomfort which could be expected from periodic vibration presentations at the location), and distraction (how likely vibrations presented at the location would disrupt mental processes associated with ongoing tasks).

Ratings varied greatly, but in terms of mean ratings, the upper arm, forearm, and back were deemed least obstructive, while the thigh and waist would be the most obstructive sites for vibration presentation (see Figure 3.4). The thigh was deemed the most annoying location for vibration presentation by far, followed by the waist. The least annoying locations were the forearm, wrist, and back (see Figure 3.5). Finally, the upper arm and back were deemed the least distracting locations for the presentation of vibrations, while the thigh and wrist were deemed to be the most distracting locations (see Figure 3.6).

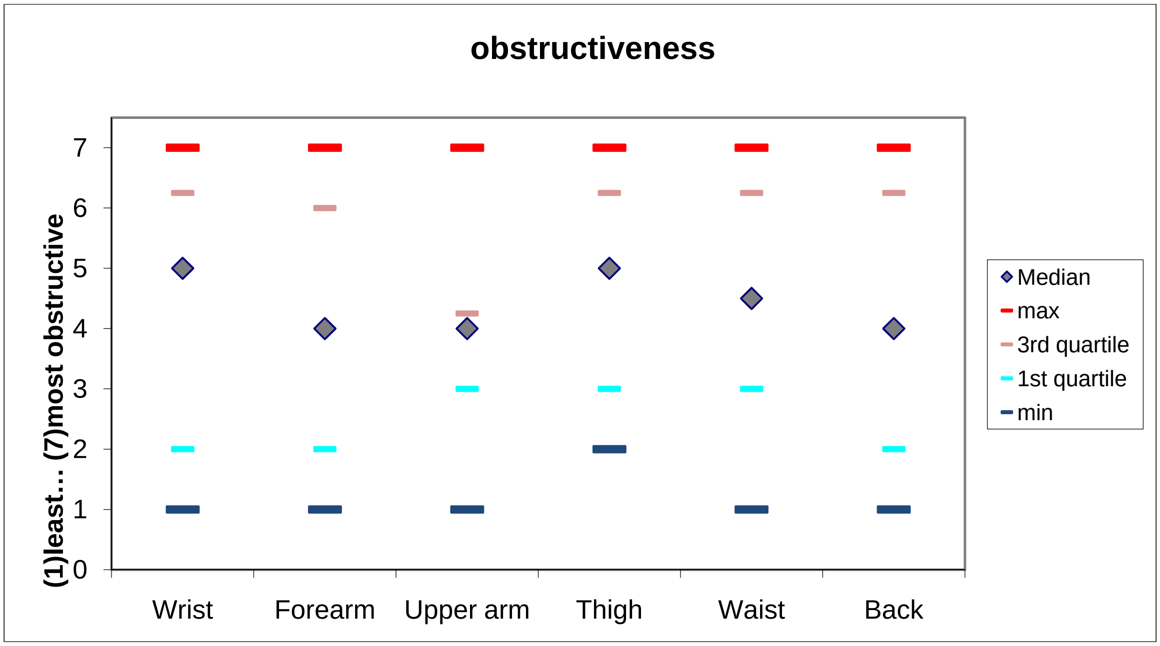


Figure 3.4: Box plots of responders' ratings of "obstructiveness" of a vibrotactile display at several body sites.

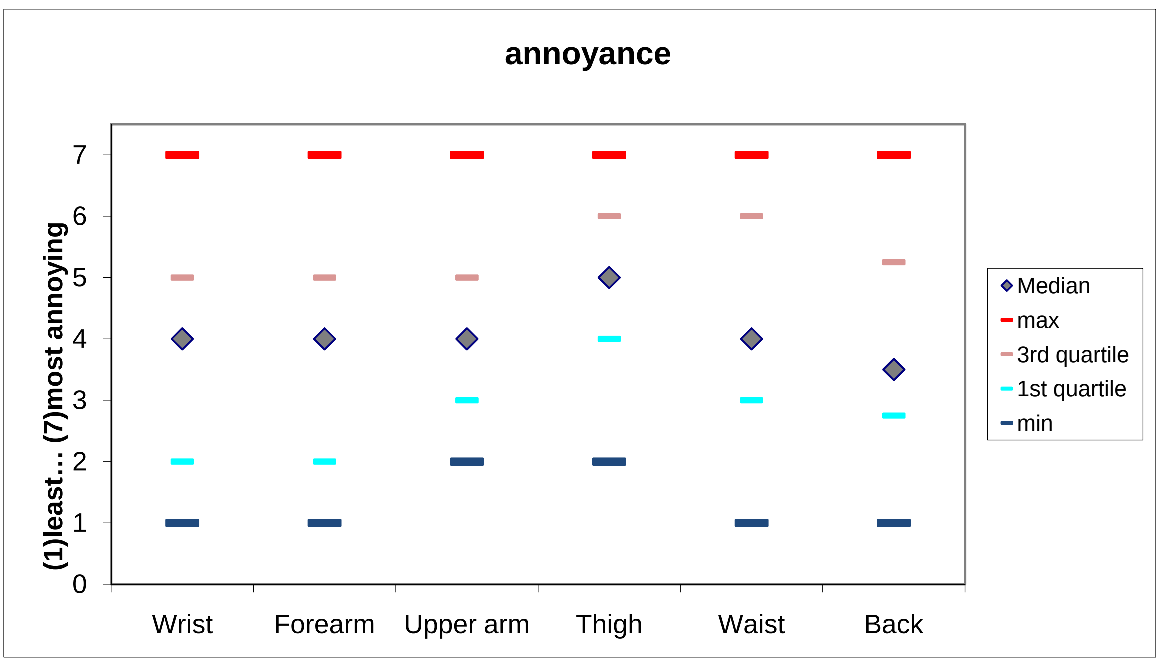


Figure 3.5: Box plots of responders' ratings of "annoyance" for vibration presentations to several body sites.

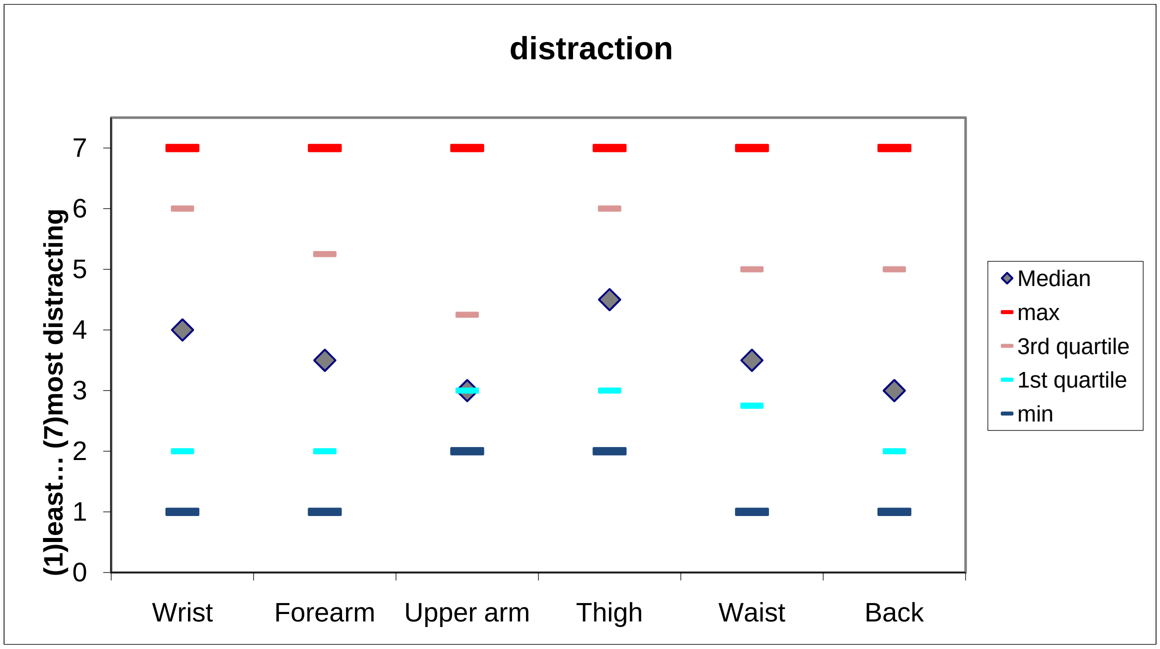


Figure 3.6: Box plots of responders' ratings of "distraction" that would result from vibration presentations at several body sites.

A sequence of four questions asked about preferences for when and how often to present, via a vibrotactile signal, the three types of information that each responder felt were most useful. These questions are listed below, with brief descriptions of general patterns in responses.

1. **Should the frequency of presentation be preset or adjustable by anesthetic personnel?**

   Across all types of information, responders highly preferred being able to adjust presentation frequency (86%). The information related to "BP drops/skyrockets" showed the highest rate of "preset" preferences, with 37%.

2. **Considering the information will be displayed via vibration patterns, should it be periodically presented throughout the case, or only during a potential BP problem?**

   Across all types of information, most responders (60%) felt that the vibration presentations should only come during blood pressure problems. Twenty-eight percent felt the information should be presented throughout the induction phase of anesthesia, if not throughout the entire surgical case. Eleven percent responded that their response depended on further information, such as specifics about the health of the patient or the surgical procedure, and “how easily the signal can be ignored when something more important is going on.” Of the individual types of information, responders most frequently preferred systolic pressures and MAP to be presented only when a problem was present, but direction and rate of change information was desired more continuously throughout the case.

3. **During such a problem, should this info be presented (as vibrations) for as long as the problem persists, or only at the start of the potential problem?**

   This question was posed only if participants responded “only during a potential BP problem” or “it depends” to the previous question. Responses to this question were fairly balanced, showing 36% preferred the vibrations to continue as long as the problem was present, 39% preferred a single vibratory notification at the beginning of a potential problem, and 26% responded that it depended on additional information. Representative “it depends” responses included that a periodic reminder would be beneficial, or additional notifications when the situation gets worse. Others mentioned a desire to be able to “silence” the vibrations, which would make it more acceptable if they persisted throughout a problem.

4. **If the display will not present this information for the entire time that the problem persists, would you prefer to have a vibratory notification once the problem is resolved?**

   This question was only posed to those who responded “only at the start of the potential problem” to the previous question. Most responses to this question were “no notification needed” (63%), but 30% would prefer to receive such a notification. The remaining responses (7%) were “it depends”, mostly listing the same reasons as those given in the previous questions.

4) Identification of natural mappings between types of blood pressure information and vibrotactile encoding mechanisms for relating the information.

For the types of information responders identified as most useful, they were asked to choose dimensions/modulations of the vibrotactile signal which most intuitively or naturally communicated the information. They were told that the vibration dimensions that could be modulated included:

- Amplitude
- Frequency
- “Intensity”: paired changes in both the frequency and amplitude of the vibration (achieved by applying higher voltages to vibrotactile devices)
- Pulse frequency/rhythm: presenting sequences of “pulses” defined by squarewave activation signals
- Body location

Figure 3.7 lists the frequency of responses for each of the top five “most useful” types of blood pressure information. Note that respondents needed to have selected a type of information as one of the three most useful to be given a chance to choose a modulation for it. Pulse frequency and intensity were chosen most often across all the types of information. Frequency was selected least often. Within each information type, responses about the most natural modulation varied considerably. However, pulse frequency was most often chosen to relay “rate of change” information, and intensity was chosen most often to communicate information related to mean and systolic pressure levels.

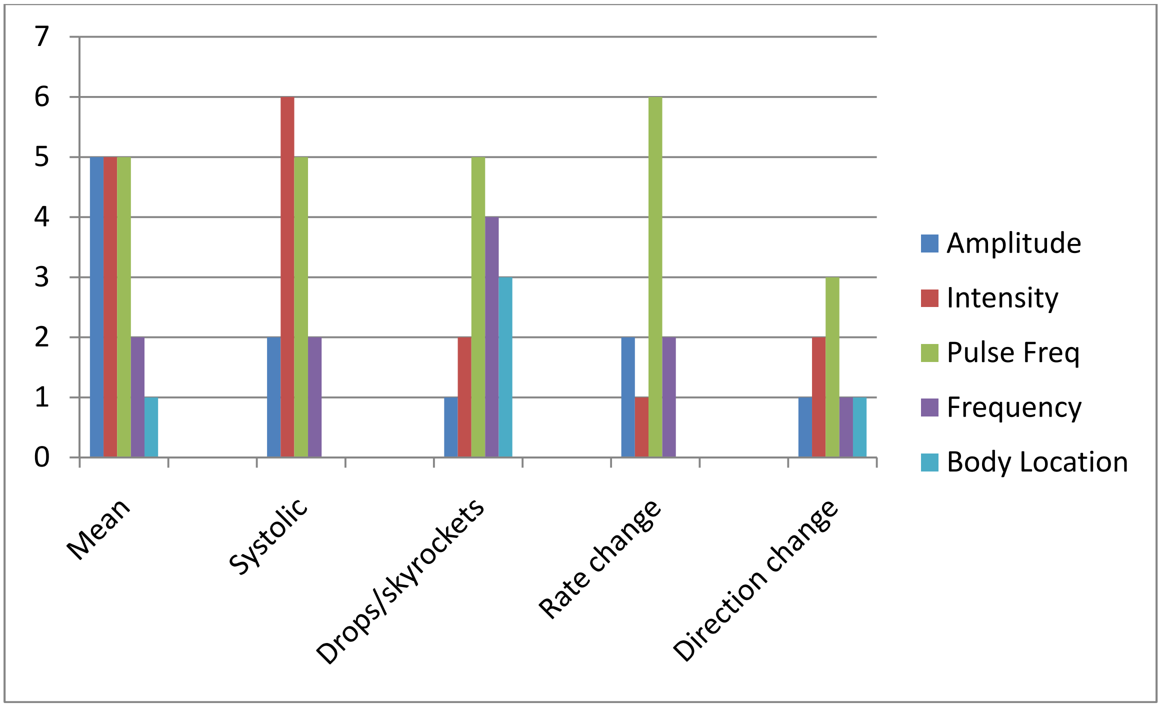


Figure 3.7: Response frequency for vibration dimensions/modulations that would most intuitively communicate each of the five most useful types of blood pressure information.

5) Input regarding physiological measures, other than blood pressure, that could benefit from improvements in display methods.

At the conclusion of the survey, respondents were given a chance, via free-response, to list any other physiological measures which might benefit from some form of vibrotactile display. Several responded that measures of capnography/end-tidal carbon dioxide (ETCO2) levels could benefit from such a display. They explained that such a display would be especially useful during intubation tasks, to aid in detecting whether the endotracheal tube had been inserted into the trachea or the esophagus. Other suggestions included a vibrotactile pulse oximetry display, a bispectral index score (BIS: a measure of brain activity that might indicate the patient's depth of anesthesia) display, a heart monitor that could present coded notifications for various arrhythmias, and a notification related to operation of infusion systems. Finally, some felt a simple vibrotactile display system could be used to present reminders for certain periodic activities, such as when antibiotics should be re-dosed.

Chapter 4

A First Investigation of the Use of Vibrotactile Displays to Support Physiological Monitoring and Multitasking: Iconic Tactile Displays

The survey described in the previous chapter served to highlight two key types of information which are considered most important in the task of blood pressure monitoring: the level of blood pressure and its direction/rate of change. It also suggested some intuitive ways to relay this information via modulations of the tactile signal. This chapter describes a study which was the first step in determining how to most effectively communicate this information via a tactile display. The study used an iconic representation (tactile icons, or “tactons”), encoding the level and rate of mean arterial pressure (MAP) into different characteristic tactile patterns, which were presented semi-continuously. It then investigated the relative performance benefits of displaying each type of information individually, and the benefit of displaying both types of information (the level and the rate), interleaved in a complex iconic signal. One of the questions this study sought to address was whether a more informative but more complex signal better supported multitasking performance (performance in a physiological monitoring task and a patient interaction task), or whether a simpler signal was more beneficial for overall performance. The findings from this study served to inform the tactification displays described in Chapter 5, in terms of the types of information to emphasize, and the degree of complexity in the signal that could support interpretation of critical states and dynamics without hurting performance on concurrent tasks.

INTRODUCTION

Tactile displays represent a promising approach to supporting the attention and task management of anesthesiologists in the clinical setting. By reducing some of the load on visual (and auditory) channels, this underutilized channel can be employed to keep anesthesiologists more continuously aware of developments in the patient's status, thus improving performance in physiological monitoring tasks, while minimally distracting them from concurrent patient care tasks. In particular, iconic vibrotactile signals called "tactons" seem worth exploring for communicating fairly rich patient health information. Tactons are structured, abstract patterns of vibrations that encode information by modulating several dimensions of the tactile signal, including the frequency, amplitude, body location, and/or waveform of the vibration (e.g., Brewster & Brown, 2004; Brown, Brewster, & Purchase, 2005). Recent studies have shown that tactile displays communicating different alarm conditions in an OR setting can be interpreted accurately (Ng, Barralon, Schwartz, Dumont, & Ansermino, 2008), and improve response times to alarming events (Ford et al., 2008). One important question is whether a more continuous tacton display of the state of and/or changes in a physiological parameter may support anesthesiologists in routine monitoring of the parameter while they are engaged in a demanding visual task. The proper design and use of such a continuous tactile display could facilitate the detection and identification of trends (e.g., slow increases in blood pressure) before they lead to alarming events.

In an effort to inform the design and use of continuous tactile displays, a recent survey of 34 practicing anesthesiologists was conducted (see Chapter 3). The survey results showed that blood pressure, specifically Mean Arterial Pressure (MAP), was both especially critical and difficult to monitor during and immediately after the "induction" phase of anesthesia. One of the tasks during this phase, endotracheal intubation (insertion of a breathing tube for surgeries involving general anesthesia), requires a good deal of focused visual attention and manual interaction and was identified as one of the most challenging tasks for patient monitoring. Therefore, the current study examined the effectiveness of various continuous tacton displays for supporting MAP monitoring and management while completing a simulated intubation task.

It was expected that the redundant display of MAP information, via tacton displays, would result in the early detection of undesirable trends and lead to fewer alarm conditions, more effective management of alarm states, and better performance on the intubation task when compared to performance with traditional MAP display methods alone. It was also expected that a tacton display which communicated information about both the state and dynamics of blood pressure would result in superior performance over the tacton displays presenting state alone or dynamics alone. In addition to providing insight into a potential means for improving monitoring performance and attention management for anesthetic personnel in the OR, the results of this study can be used to inform the design of tactile displays for other complex multitask environments, such as car cockpits and flight decks.

METHOD

Twenty students from the University of Michigan played the role of anesthesiologist and were responsible for two concurrent tasks: endotracheal intubation of a simulated patient, and a physiological monitoring task.

Endotracheal intubation simulator and task

The process of endotracheal intubation involves a series of steps which, depending on the skill of the anesthesiologist and the health and anatomy of the patient, can take anywhere from less than a minute to several minutes. For the purposes of this study, a simulator was developed that required activities related to the pre-intubation ventilation (administering pure oxygen for 1 minute to prepare the patient for the intubation process), the intubation itself (insertion of the laryngoscope, displacement of the tongue to visualize the laryngeal opening, insertion and inflation of the tube), and verification of the tube placement (using a stethoscope to auscultate each lung to verify breath sounds and the stomach to verify their absence). By simulating vomit in the airway and incorrect placement of the tube after the first intubation, the entire procedure had to be started over from the beginning twice, and thus each experimental scenario involved essentially three iterations of the intubation process. In all, completion of the entire intubation scenario took between approximately 7.5 and 10 minutes.

A 30-inch desktop monitor displayed a 3D patient model in a simulated OR environment. The model included all the relevant anatomy that would be visible in the process of intubation (see Figure 4.1), as well as the patient's head and torso for pre-ventilation and auscultation. A continuously-playing sound clip of ambient OR sounds served to improve the realism of the simulator and mask the sound of tactor activation.

Participants performed the intubation procedure with two controllers. The primary controller was a wireless gyroscopic remote developed for the Nintendo® Wii™ gaming system, which was configured to allow participants to pick up and manipulate tools (e.g., laryngoscope, ventilation bag, stethoscope) in 3D space. This control style allowed natural mapping of movements, such as the insertion of the laryngoscope blade. The secondary desktop mouse controller was used for tasks which involved both hands (e.g., inserting the tube while holding the laryngoscope in the proper position).

Performance on the intubation task was measured as the time to complete the scenario (if within 10 minutes), or a projected estimate of completion time based on the percentage of the scenario completed at 10 minutes. If participants finished intubating within 10-minutes, they were instructed to then begin a data entry task in which they transcribed a list of drug dosages into a patient log. This assured competition for visual and manual resources throughout the duration of the 10-minute monitoring task which was performed concurrently.

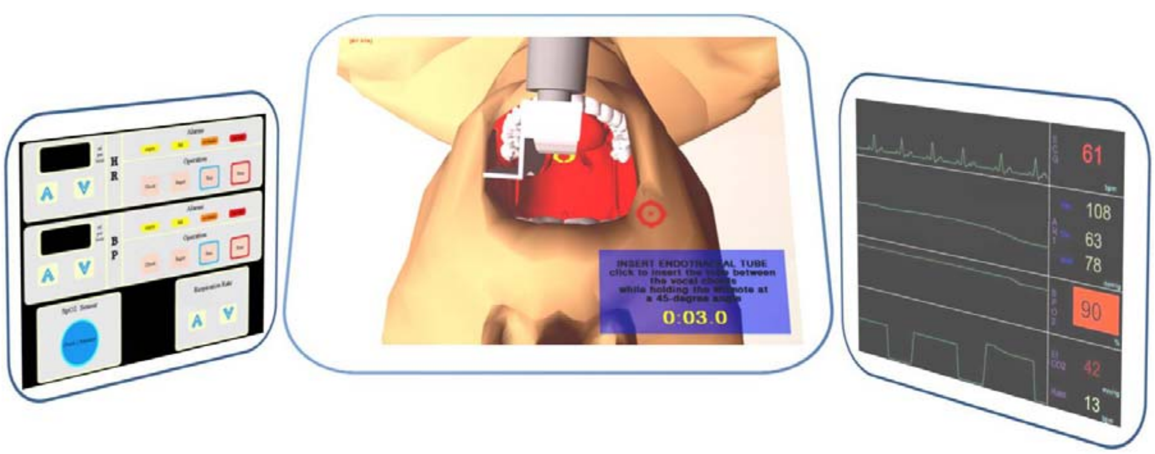


Figure 4.1: The simulation setup, distributed across three separate desktop monitors. Left to right: the physiological parameter controllers (touch screen), the intubation simulation display, and the physiological parameters display.

Physiological parameter display and monitoring task

Four critical physiological parameters were displayed on a desktop monitor located to the right of the intubation screen (see Figure 4.1): 1) heart rate (ECG), 2) three components of blood pressure as if measured by an intravenous pressure sensor – Systolic, Diastolic, and Mean Arterial Pressure (MAP), 3) oxygen saturation in the blood (SpO2), and 4) measures of CO2 in exhaled gases (EtCO2) and respiration rate. Adjacent to each parameter was a scrolling history graph that displayed the characteristic ECG and EtCO2 waveforms as well as history values for MAP and SpO2 (see Figure 4.2). Alarming states for each parameter were communicated visually by backlighting the value, as shown with SpO2 in Figure 4.2.

As in most OR environments, each heartbeat was announced with an auditory beep in addition to the visual display, with the pitch of the beep mapped to the SpO2 value (higher pitch for higher SpO2%). Each parameter also had an associated auditory alarm, which accompanied the highlighting of the parameter when it was outside of alarm thresholds.

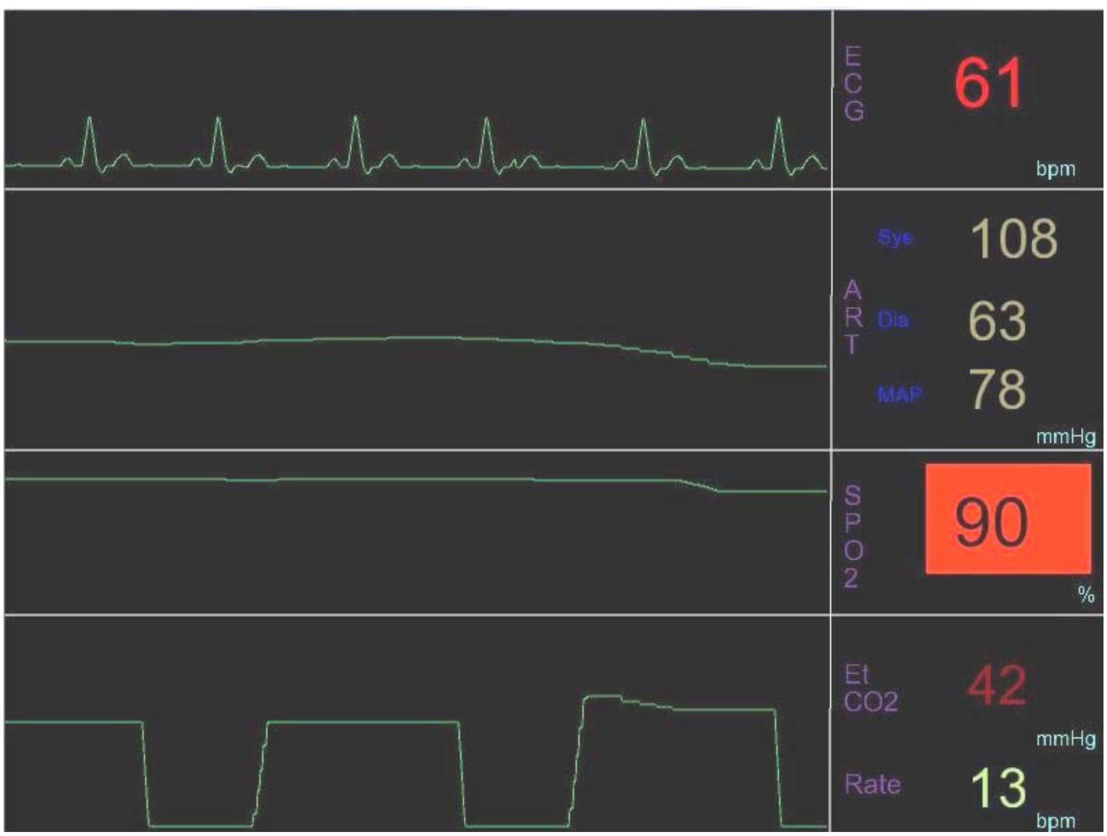


Figure 4.2: Desktop display of physiological parameters.

The numerical values of each parameter were updated once per second, and a 10-minute simulation script was used as input for fluctuations of each parameter. By controlling the drug infusion levels and respiration rate and "replacing" the SpO2 sensor, participants were expected to maintain all parameters within alarm thresholds as much as possible for the duration of the 10-minute scenario. Without such corrective actions, the MAP parameter was scripted to exceed alarm thresholds (and activate an alarm) six times, and each other parameter was scripted to do so twice. Participants were instructed that MAP was an especially critical parameter for this patient, and that his preexisting vascular disease would cause frequent and sometimes rapid changes in this parameter.

Performance on the monitoring task was measured for each experimental condition in several ways: 1) the number of MAP alarms (which occurred each time levels dropped to 70 mmHg or below, or rose to 100 mmHg or more); 2) the percentage of scenario time that MAP levels were in alarming ranges; 3) the sum of how far alarm thresholds were exceeded at each second that MAP was at an alarm level; 4) the number of alarms for all other physiological parameters; and 5) the percentage of scenario time these other physiological parameters were at alarming levels. Because each alarm state involved a rather salient auditory signal, pure response times to the alarms were not expected to differ between conditions and thus were not measured. Instead, the measures emphasized alarm avoidance and management, i.e., returning alarming parameters to a safe state, thus giving an advantage to being better informed about the dynamics of the parameter.

Tacton displays

Tacton displays were used to redundantly communicate information regarding the state and dynamics of MAP. The tacton signals were presented via two vibrating "tactor" devices (C-2 tactors developed by Engineering Acoustics, Inc.). These tactors were arranged vertically, secured via an elastic belt to the mid-lower back and via shoulder strap near the shoulder blade, offset approximately 10 cm to the right of the spine (on the same side as the physiology display). The back was chosen as the presentation site as it has proven effective in earlier studies (e.g., Chapter 2: Part B) and was identified by anesthesia providers as a location that was unlikely to interfere with existing tasks (Chapter 3). The tactors oscillated a 7.5-mm "skin contactor" at frequencies near 250 Hz with a displacement of approximately 1 mm.

Anesthesia providers identified the level (current reading) and direction/rate of change of MAP as some of the most critical information during and immediately after induction (Ferris, 2008). Therefore three tactile patterns were developed: 1) the *level* tacton communicated the current approximate level of MAP; 2) the *rate* tacton communicated the direction and rate of change of MAP as measured over a moving 3-second window; and 3) the *level+rate* tacton communicated both level and rate of change combined in one complex pattern.

The *level* tacton involved two successive 0.25 s vibration pulses presented every 2 s from one of the tactors when the level was outside of the range between 80 and 90 mmHg, and would cease when within this range. Levels below 80 would vibrate the bottom tactor, and those above 90 would vibrate the top tactor. By communicating that MAP was outside of a relatively centered range, and additionally using body location to encode whether it was closer to the lower (70 mmHg) or higher (100) alarm threshold, it qualified as a tacton display.

The *rate* tacton involved 0.75 s vibrations presented every 2 s from the top (increasing MAP) or bottom (decreasing) tactors, notifying participants if MAP dramatically rose or fell in a short period of time. If the level did not change, or changed minimally (2 mmHg or less) over the previous 3 s, no vibration was presented. The magnitude of the change was communicated by amplitude modulation (e.g., Brown et al., 2005) – mathematically combining signal frequencies that are near 250 Hz so that the dissonance is perceived as a sinusoidal pulsation of amplitude peaks commonly referred to as "beats". The resultant beat frequencies were 2, 3, 4, and 6 Hz, communicating larger rates of change with higher frequencies.

The combined *level+rate* tacton was identical to the *level* tacton when MAP was outside of the 80 - 90 range and stable (not changing dramatically). When outside this range, if MAP continued toward the alarm threshold, the longer and distinctly different rate pattern would replace the second of the two vibration pulses for the *level* pattern, indicating how quickly MAP was approaching an alarm state.

Procedure

After familiarization with the simulation and controllers, a 45-minute training session introduced participants to the two tasks and three tacton displays. Training concluded with participants completing 4 shortened segments of the intubation task while maintaining all physiological parameters within alarm limits for the duration of the segment. These segments allowed practice with each experimental condition – a baseline condition and three employing the different tacton displays. All participants were able to complete the training successfully, after which they completed the four 10-minute experimental conditions. To counteract learning effects, the order of the experimental conditions, and the simulation scripts associated with each condition were balanced between participants.

RESULTS

Data from two participants were removed from analysis after one experienced an apparent hardware failure and one's performance measures were outside of three standard deviations for most of the metrics. The data from the remaining 18 participants (8 females, average age: 23, sd: 3.5) were analyzed in repeated measures general linear models (formulated in SPSS 16.0). Performance was compared between conditions for both the intubation and monitoring tasks.

Performance in the intubation task

Likely due to large interindividual performance differences (note the large standard deviations), no significant effect was found for experimental condition on time to complete the intubation procedure. However, trends suggested that the times were longer in the baseline condition (average completion time: 856.3 s; SD = 643.4 s) than in the *level* (average = 659.8 s; SD = 286.4 s; $p = .082$), *rate* (621.9; 129.4; $p = .104$), and *level+rate* (729.1; 488.7; $p = .220$) conditions.

Performance in monitoring and managing MAP

The average number of alarms specifically for MAP (when MAP was outside of the alarm thresholds of 70 and 100 mmHg) differed between conditions ($F(3,15) = 6.636$; $p = .005$; see Figure 4.3). Post-hoc comparisons showed that the baseline condition (6.9 alarms) was significantly higher than the *level* (5.2 alarms; $p = .014$), and *level+rate* (4.2; $p < .001$) conditions, but not significantly different from the *rate* condition (5.9 alarms). The *level+rate* tacton also showed a significant advantage over the *rate* tacton ($p = .009$) and a trend suggesting an advantage over the *level* tacton ($p = .155$).

The percentage of scenario time during which MAP was at alarming levels was also significantly affected by condition ($F(3,15) = 11.608$; $p < .001$; see Figure 4.3). This percentage was significantly higher in the baseline condition (8.9%) than all three tacton conditions – *level* (4.0%; $p < .001$), *rate* (7.1%; $p = .049$), and *level+rate* (3.7%; $p <.001$). The *rate* tacton also resulted in significantly higher percentages than *level* ($p = .014$) and *level+rate* ($p < .001$) tactons.

For each second that MAP levels were in alarm states, the sum of the magnitudes outside of the alarm threshold limits (i.e., the area between curves of alarm thresholds and MAP value plotted over time) was similarly affected by experimental condition ($F(3,15) = 6.812$; $p = .004$; see Figure 4.3). The average sum of 226.9 mmHg-s for the baseline condition was considerably larger than those for the *level* (81.8 mmHg-s; $p = .002$) and *level+rate* (72.4; $p < .001$) conditions. The *level+rate* tacton also resulted in significantly better performance than did the *rate* tacton (203.6 mmHg-s; $p = .025$).

Performance in managing other (non-MAP) parameters

Experimental display condition had no significant effect on the number of alarms for the heart rate, SpO_2, and $EtCO_2$ parameters, nor on the percentage of scenario time in which these parameters were in alarm states. The average number of non-MAP alarms per scenario was 8.3 in baseline, 8.8 in *level*, 8.6 in *rate*, and 8.7 in *level+rate* conditions. The average percentage time spent with these parameters in alarm states was 11.9% in baseline, 11.5% in *level*, 11.1% in *rate* and 13.4% in *level+rate* conditions.

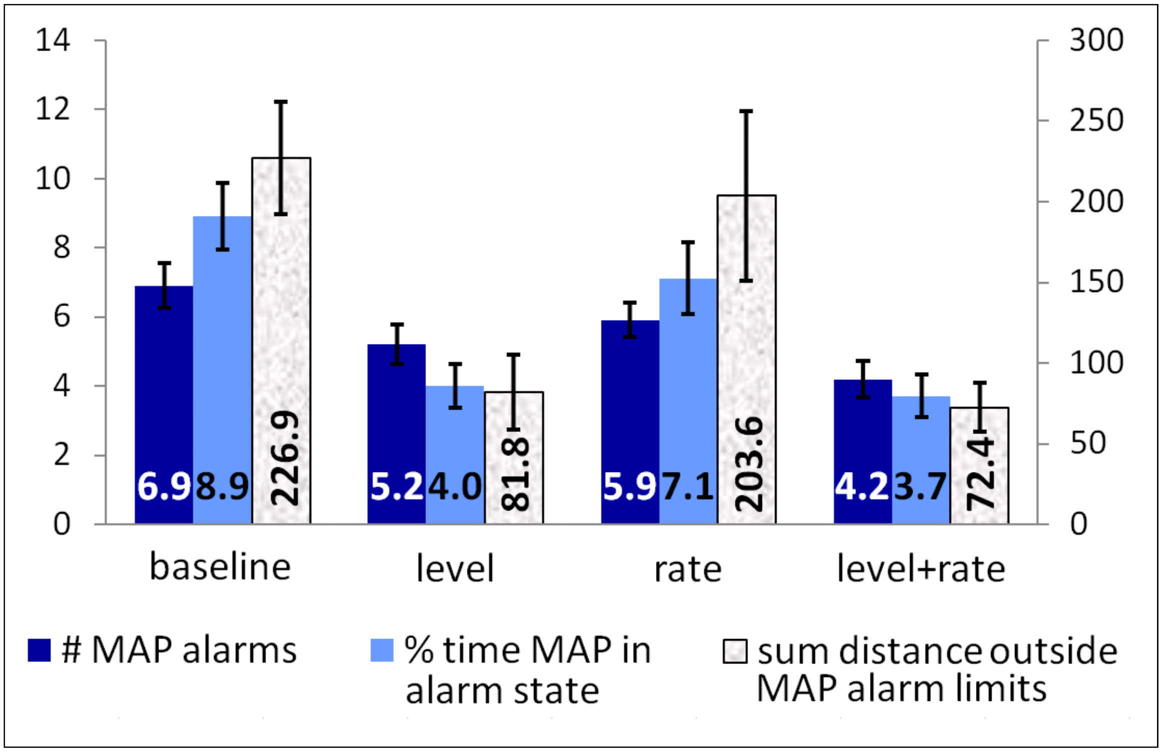


Figure 4.3: Results related to MAP management. Associated with the left axis: number of MAP alarms (dark blue bars) and percentage of scenario time that MAP was in an alarm state (light blue bars). Associated with the right axis: sum of distances outside of alarm limits (light grey textured bars). Error bars represent standard error.

DISCUSSION

The goal of this study was to determine the potential performance benefit of iconic tactile displays for supporting continuous anesthetic monitoring. These displays can help offload vision and hearing and better support the early detection and avoidance of dangerous trends in physiological state. In particular, the experiment investigated the relative effectiveness of different tactile patterns that presented information on the level of MAP, its rate of change, or both types of information.

The results of this study provide evidence that tactile displays can improve anesthetic monitoring while anesthetic personnel are engaged in demanding visual tasks, such as endotracheal intubation. Significant improvements in monitoring performance were observed when tactons redundantly presented patient MAP data, resulting in fewer MAP alarms, less time in an alarming state, and smaller exceedance of alarm thresholds. These improvements may be attributed to being better informed of MAP state and, importantly, dynamics. Rather than waiting until an alarm state, detecting and treating potentially dangerous trends early reduces the risk of adverse consequences to patient health. For example, one reason anesthetic personnel are concerned with MAP dynamics during patient induction is that steady or sudden drops in blood pressure can indicate an overdose of the gases used to induce anesthesia (Chapter 3). These drops are not easy to detect visually and may not immediately lead to MAP levels low enough to activate threshold alarms. When alarms do sound it may mean the patient has already overdosed, which could be life-threatening and/or have lasting negative health effects.

While the *rate* tactons improved only the time that MAP was in an alarming state, the *level* and *level+rate* displays showed dramatic improvement in all three performance measures for MAP monitoring. These performance differences may reflect the relative usefulness of the information being presented for MAP management, the ease with which modulations of the tactile signal could be interpreted, or a combination of these factors. Some participants reported that the erratic nature of the MAP led to the *rate* tactons being presented too frequently to be very meaningful, despite the relative infrequency of signals that communicated changes that were large in magnitude. These reports may also reflect a greater difficulty interpreting the encoding method for the magnitude of change – amplitude modulation – than the level encoding method (body location). Similar findings from past studies support this explanation (e.g., Brown et al., 2005).

It is interesting to note that MAP monitoring performance with the *level+rate* tactons was consistently the best, showing significant improvements over the *rate* tactons, and a trend favoring improvement over the *level* pattern. This suggests that participants were able to interpret the complex pattern, and thus were better informed and could most effectively manage MAP. It also demonstrates the value of rate information when level information is also considered, as the combined pattern can directly map to the urgency of attending to MAP (e.g. high *and increasing*, or low *and decreasing* MAP).

A trend suggested improved performance on the intubation task in the three tacton display conditions over the baseline condition, which likely reflects a reduced competition for visual resources. While traditional physiological displays require frequent disruption of the intubation task to visually sample relevant parameters which delays completion of the procedure, the ability to be informed of these parameters via non-visual means reduces the need for frequent re-orienting of attention. It was interesting to note that for the intubation task, a trend suggested performance with the tactons which included the most information (*level+rate*) may have been worse than with the simpler tactons. This might suggest that while the more informative signals might better support physiological monitoring, they might also impose a higher workload which manifests itself as performance decrements in the concurrent intubation task.

The apparent non-effect of display type on performance in monitoring the other (non-MAP) parameters is not surprising, since the tactons redundantly communicated only information related to MAP. The fact that performance in monitoring these other parameters did not get worse shows that improved MAP monitoring was not due to a tradeoff in monitoring emphasis.

Rather large individual differences in multitask performance were observed, which are reflected in the large standard deviations for some measures. While most participants found the tacton displays helpful and, once trained, could use them effectively for nonvisual monitoring of MAP, a few participants reportedly were, at times, overwhelmed by the simultaneous visual, auditory, and tactile input, to the point that they "tuned out" the tactile signals. This was reported most often with the most complex tactons, the *level+rate* presentations. These reports may reflect processing interference related to the particular encoding methods used to define the displays (i.e., whether information is coded spatially or temporally/categorically; Wickens, 2002). For example, MAP rate of change was encoded temporally, so its interpretation may interfere with auditorily monitoring heart rate or compressing the simulated ventilation bag at regular intervals – concurrent tasks which also involve temporal processing. If this processing interference leads people to "tune out" the tactons, then presenting the information via different modulations of the tactile signal might lead to more universally interpretable, and thus useful tactile displays (e.g., Chapter 2: Part B). One way to ensure that this processing interference is minimized, for example, is to redundantly encode information via separate methods, each of which requires a separate processing code. This would allow engaging either processing code – whichever faced less interference – in interpreting the tactile message.

The results from this study demonstrate the potential for tactile displays to better support non-visual anesthetic monitoring and management of patient physiology. By keeping anesthesiologists better informed of the state and dynamics of critical parameters without imposing further load on visual and auditory resources, monitoring breakdowns, and critical incidents that result from breakdowns, can be reduced. The final study described in Chapter 5 builds on these results by using a more comprehensive theory-based approach to the design of the displays. This included designing signals that were less likely to be affected by interference at the perceptual (e.g., Chapter 2: Part A) and cognitive (e.g., Chapter 3: Part B) processing stages. These designs also investigated the benefits of a more natural way to relate some of the more complicated signals, in particular, the rate of change information. This was necessary because the tactification displays represented significantly more complex displays, expanding on the tacton displays to include respiration measures as well as MAP information. Finally, the evaluation study described in Chapter 5 investigated the benefits of each tactification display with practicing anesthesiologists as participants (rather than engineering students, as this study employed), and was set in a much more complex simulation environment that more accurately reflected the displays and tasks anesthesiologists face in the clinical setting.

REFERENCES

Brewster, S. & Brown, L.M. (2004). Tactons: Structured tactile messages for non-visual information display. *Proc 5th Conference on Australasian User Interface*, 15-23.

Brown, L.M., Brewster, S.A. & Purchase, H.C. (2005). A first investigation into the effectiveness of tactons. *Proc World Haptics '05*, 167-176.

Ford, S., Daniels, J., Lim, J., Koval, V., Dumont, G., Schwarz, S.K.W., & Ansermino, J.M. (2008). A novel vibrotactile display to improve the performance of anesthesiologists in a simulated critical incident. *Technology, Computing, and Simulation, 106*(4), 1182-1188.

Ng, G., Barralon, P., Schwartz, S.K.W., Dumont, G., & Ansermino, J.M. (2008). Evaluation of a tactile display around the waist for physiological monitoring under different clinical workload conditions. *30th Annl Meeting IEEE Engineering in Medicine and Biology Society*, 1288-1291.

Sklar, A.E., & Sarter, N.B. (1999). Good vibrations: Tactile feedback in support of attention allocation and human-automation coordination in event-driven domains. *Human Factors*, *41*(4), 543-552.

Wickens, C.D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Sciences*, 3, 159-177.

Chapter 5

Design and Evaluation of Tactification Displays

As outlined in Chapter 1, the ultimate goal of the research described in this dissertation is the development of a "tactification" display – a novel type of vibrotactile display that can keep an operator continuously informed of the state and dynamics of a monitored system (in this case: to support the attention and multitask performance of anesthesiologists in a clinical setting). Several research activities, described in Chapters 2 - 4, contributed input to the final designs for these displays. This chapter describes three tactile displays, two of which could be considered "tactifications", and one which is more of an "informative alarm" system. The design of each of these displays was motivated by theory-based and practical considerations to support an anesthesiologist in a multitask set that is representative of their responsibilities in the clinical setting.

The first part of this chapter describes the design considerations and how these led to the design of the three tactile displays. Next, a set of hypotheses are postulated that predict how the displays will affect the performance and workload of anesthesiologists. These hypotheses relate to the 6 goals for improving physiological display technologies described in Chapter 1 (pages 8 and 20-23). Finally, an evaluation study designed to test these hypotheses is described in detail. The study was conducted with anesthesiologists in a complex simulation environment. Its results are reported and discussed individually, followed by a general discussion of the most important findings and their implications for design.

TACTILE DISPLAYS

Chapter 4 showed how tactile presentation of data related to patient blood pressure (as continuously measured via an intra-arterial catheter, colloquially referred to as an “a-line”) can improve performance with regard to managing that specific parameter. Following that success, the final tactification displays described in this section were also designed to relate a-line blood pressure data (Mean Arterial Pressure over the period of each heartbeat, or MAP), but additionally to relate data for respiratory measures of exhaled partial pressures of carbon dioxide (end-tidal CO2, or ETCO2), tidal volumes (TV; the volume of gas that is inhaled/exhaled with each breath), and respiration rate (RR; the number of breaths per minute). From a practical perspective, this makes the displays more useful in detecting and diagnosing critical health states. Following pulse oximetry, capnography (i.e., the continuous display of ETCO2 levels) and arterial blood pressure displays are cited as some of the most useful monitors for detecting and diagnosing events that could lead to adverse patient outcomes (James, 2003; Rady, Ryan, & Starr, 1998; Webb et al., 1993). A display that combines capnographic measures with tidal volumes and respiration rate may in fact be used to detect critical events more often than any other single monitor (Watson & Sanderson, 2004; Webb et al., 1993). “Capnography” was also listed frequently when participants were asked in a survey (Chapter 3) which patient health parameters, in addition to blood pressure, could be displayed via a tactile display to benefit anesthetic monitoring performance.

The concurrent display of four parameters also allows the tactification displays to be compared to sonification displays of patient health data that are prevalent in the literature and commonly include five or more variables (e.g., Fitch & Kramer, 1994; Loeb & Fitch, 2002; Sanderson et al., 2008; Seagull, Wickens, & Loeb, 2001; Watson & Sanderson, 2004). Note that these sonification displays include blood oxygen saturation and heart rate as two of the sonified variables. As these measures are currently displayed very effectively in clinical settings via the pulse oximetry sonification, neither was included in the tactile display designs. The evaluation study for the tactile displays did, however, include a requirement to monitor the auditory pulse oximetry display in addition to the tactified variables, therefore representing a similar monitoring load to that in the sonification studies.

The tactile displays consisted of vibrotactile devices ("tactors") secured to the upper left arm (blood pressure measures) and back (respiration measures). One reason these body locations were chosen as presentation sites was anesthesiologists' low subjective ratings for obtrusiveness (while conducting a clinical task set), annoyance, and distraction in survey responses (Chapter 3). Another reason was to support natural mapping between display location and the underlying data (see the Design Considerations section below).

The vibrotactile devices employed for these displays (C-2 tactors developed by Engineering Acoustics, Inc.; http://www.eaiinfo.com/EAI/Tactor%20Products.htm) were secured to participants' upper left arm and back with a snug but comfortably-fitted garment. Each tactor (see Figure 5.1) oscillates a 7.5-mm "skin contactor" at 250 Hz with intensities on the order of 20 dB above the sensory threshold (maximum skin contactor displacement of 1 mm).

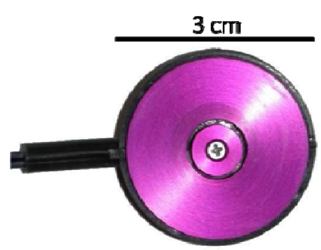


Figure 5.1: The C-2 tactor developed by Engineering Acoustics, Inc.

The tactors were attached to experimental garments with Velcro, which allowed them to be easily transitioned between different garments to accommodate each sex and differences in body size. The experimental garments chosen were medical "compression" garments, designed to maintain a consistent pressure over maximum surface area on the torso in order to aid in recovery following major weight loss or body-reshaping surgery. This property makes them ideal for presenting vibrations over the upper body surface, as it assures that the distributed tactors remain in contact with the skin at a relatively consistent pressure. Two men's garments, sized Large and Extra Large (Design Veronique Zippered Compression Vest with Arms: http://designveronique.com/cgi-bin/ic/dv2/642.html; see Figure 5.2), and a women's garment, sized Medium (Rainey Long-Sleeve Bolero: http://www.raineywear.com/shouldersbiceps.aspx) adequately accommodated the body shapes and sizes for all of the participants in the evaluation study.

Figure 5.2: Male compression garment used to secure tactors to the back and upper arm (Design Veronique).

Figure 5.3 shows the arrangement of the tactors on the men's experimental garment. Respiration measures were represented on the back, where each tactor was spatially separated by at least 65 mm, which is greater than the mean spatial localization error of roughly 11-12 mm for the back, as measured by Weber's error localization protocol (Greenspan & Bolanowski, 1996). At times, vibrations would be presented from multiple locations simultaneously on the back. The minimum distance between activated tactors in these cases was approximately 170 mm, well above the distance defined by Weber's two-point threshold protocol of approximately 40 mm (e.g., the average human can recognize multiple presentation sites when they are separated by at least this distance) (Greenspan & Bolanowski, 1996). Though not clearly visible in Figure 5.3, five tactors were also arranged longitudinally along the lateral surface of the upper left arm (right side of Figure 5.3) to represent blood pressure data. For some participants this arrangement overlapped part of the soft tissue just below the shoulder joint. The minimum distances between these tactors was 45 mm, well above the mean spatial localization error of 9-11 mm for the shoulder and upper arm (Greenspan & Bolanowski, 1996). Vibrations were never presented from more than one location at a time on the arm.

Figure 5.3: Experimental garment with tactors attached.

DESIGN CONSIDERATIONS

The design of the three tactile displays described in this chapter was driven by a number of practical and theory-based considerations. First, the decision of which patient vital information to represent in the displays – mean arterial blood pressure (MAP), end-tidal carbon dioxide partial pressure (ETCO2), respiration tidal volumes (TV), and respiration rate (RR) – was driven by literature review and input gathered from interviews and surveys. As previously mentioned, one or more of these four parameters are involved in a major portion of adverse patient outcomes (James, 2003; Rady, Ryan, & Starr, 1998; Webb et al., 1993), and a need for more effective means of displaying these data motivated much of the sonification work outlined in Chapter 1 (e.g., Fitch & Kramer, 1994; Loeb & Fitch, 2002; Sanderson et al., 2008; Seagull, Wickens, & Loeb, 2001; Watson & Sanderson, 2004). Interviews with anesthesiologists in the early stages of this research identified blood pressure as a parameter that posed one of the greatest challenges to effective physiological management under the current display configurations. The results of the survey described in Chapter 3 confirmed that an international contingency of anesthesiologists felt that blood pressure data, and in particular the MAP measure, were both critically important and insufficiently displayed under current methods. This survey also helped to identify ETCO2 as another measure that would benefit from more effective display technologies.

Next, the decision of **how** to present that information was driven, in part, by the goal to support important aspects of preattentive reference (Woods, 1995; see Chapter 1). To that end, the following design principles/considerations were taken into account: 1) employ natural mappings between the vibrotactile signals and the represented data, 2) minimize perceptual and cognitive interference, and 3) facilitate a transition from attentional "background" to "foreground".

Natural mapping

The tactile displays were designed to preserve a natural mapping between the physical properties of the display signals and the represented data, which, in turn, facilitates accurate and efficient interpretation of the data (e.g., Norman, 2002). One way this was achieved was to present the tactile signals in body locations that are likely associated with the parameter they represent. When clinicians think of blood pressure, for example, they might naturally associate this measure with the blood pressure cuff being used on a patient's upper arm. To exploit this natural association, MAP data was relayed via vibrations on the upper left arm. Similarly, the act of breathing can be naturally associated with expansion of the upper torso, where the lungs reside. The measures related to each breath (ETCO2 and TV) were therefore presented on the back, the most accessible body surface on the upper torso. The two respiration measures were further distinguished in a natural way through their associations with separate locations on the back. TV level, a measure of how inflated the lungs become with each breath, was relayed via tactors that were spatially arranged to resemble the two lungs (see Figure 5.4). ETCO2 is measured during the exhalation portion of each breath, which is characterized by gases flowing from the lungs and out of the mouth/nose via the trachea. Therefore, the tactors associated with the ETCO2 display were arranged in an axis that was in line with the trachea, bisecting the two lungs (see Figure 5.4).

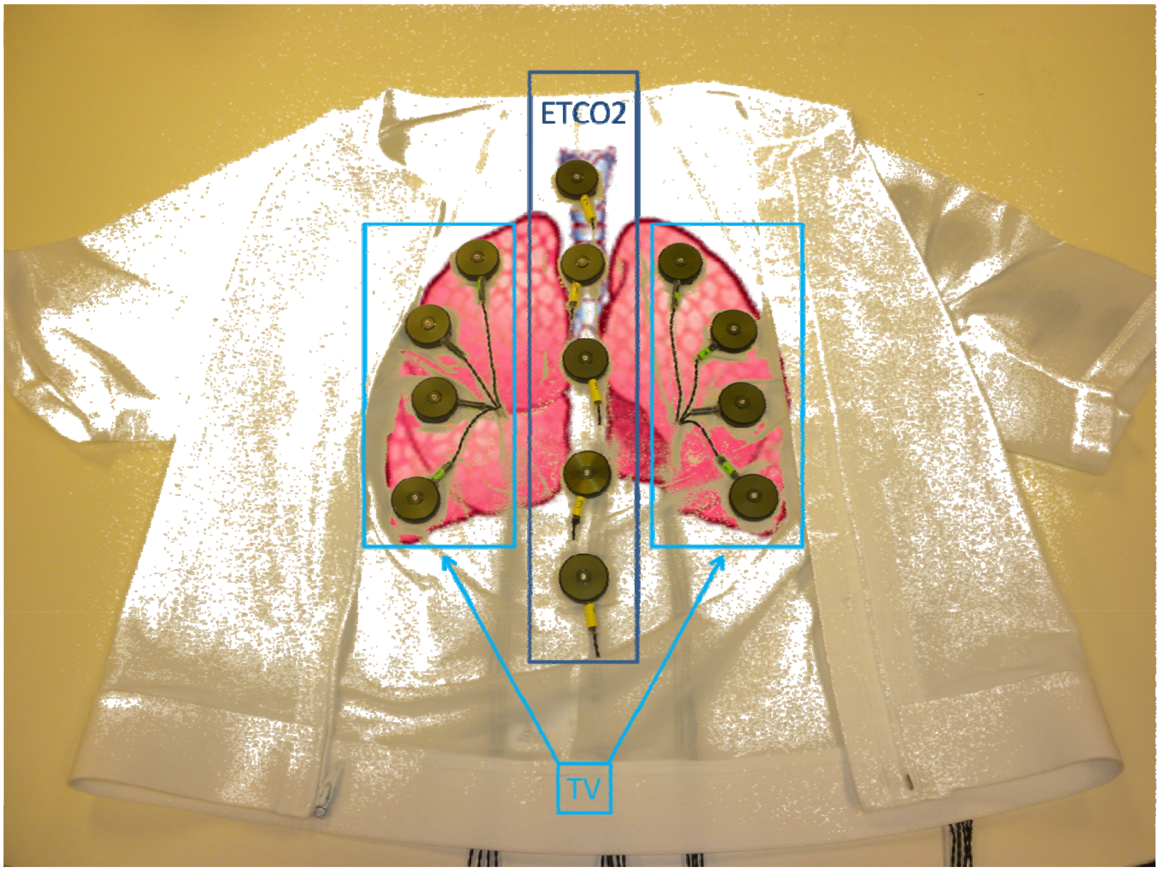


Figure 5.4: The spatial arrangement of the TV display was designed to feel like two lungs "filling" with each breath. Measures of ETCO2 were related via an axis of tactors in line with the trachea, which, as the passage for gases during exhalation, mapped naturally to that measure.

As often as possible, natural signal mappings took advantage of the spatial and temporal discrimination abilities of the sense of touch (e.g., Geldard, 1960; Welch & Warren, 1986). Because each of the parameters MAP, ETCO2, and TV are generally thought of in terms of a higher or lower number, additive representations were chosen where "high" and "low" levels were mapped to higher and lower locations on the body. As MAP/ETCO2 levels went up, vibrations would be presented via tactors higher up on the arm/back. Similarly, the filling of the lungs for each breath was represented by a sequence of vibrations from the bottom of the tactor-lungs upward. Higher TV levels caused this sequence to travel higher up the back.

The temporal properties of the displays were also used to facilitate natural interpretation of the vibrations. The tactile display of each breath was characterized by a sequence that first represented the inhalation portion of the breath (the filling of the lungs, relating TV levels), and then the exhalation portion (vibrations along the “trachea” axis, relating ETCO2 levels). The fact that TV presentations always preceded ETCO2 presentations made the vibrations related to each parameter easily distinguishable and naturally identifiable. The two tactification displays presented vibration patterns representing inhalation, then exhalation for each individual breath. Therefore, with these displays, respiration rate (RR) could also be inferred in a natural way. Abnormal rates could be recognized by the frequency of breath presentations and/or the speed with which each individual inhalation/exhalation cycle completed. For example, an increase in RR caused a resultant sensation of faster “filling” of the lungs and a shorter exhalation, and also of entire breaths occurring with increasing frequency.

The rate of change in levels of a physiological parameter is an important piece of information for diagnosing problems related to the parameter. In the survey described in Chapter 3, participants indicated that the most natural way to relate the rate of change in blood pressure would be via “pulse frequency”. This suggestion was adopted for the study described in Chapter 4; however, this led to fairly poor performance which may be explained by the unnatural requirement for rate to be presented as part of a tactile icon: information about changing MAP levels gathered over time needed to be abstracted into a single presentation that was considerably shorter than the time window it represented. In contrast, the tactification displays allowed for a much more natural way to represent the rate of change in data. Simply by communicating the approximate levels of a parameter in realtime, the rate of change for that parameter can be inferred by the magnitude of level changes and the time window over which the changes take place. In this way, the tactification displays took advantage of both the spatial and temporal discrimination abilities of touch to naturally relate information that also has both spatial and temporal qualities.

The waveform of a vibration signal was also modulated to support natural mappings. For example, the majority of respondents to the survey in Chapter 3 felt that the intensity/amplitude dimension of the vibration signal could most effectively relate blood pressure levels; this is likely because “pressure” maps fairly naturally to vibration intensity, which relates to the amount of pressure exerted by tactile devices on the skin. Thus, in addition to spatial location, vibration intensity was used to redundantly represent MAP levels on the arm. Intensity was similarly used to redundantly relate TV levels along with spatial location. This mapping is natural if one considers that as the lungs inflate, the pressures exerted by the inhaled gases on the walls of the lungs increase.

Finally, natural mapping was supported by creating metaphorically-derived vibrotactile icons (e.g., MacLean, 2008a; 2008b) to further help the receiver parse the separate components of the tactification displays. Despite occurring in separate (and naturally-mapping) locations on the body, the display components related to MAP and ETCO2 were quite similar: each involved 5 tactors, arranged longitudinally, and mapped up and down to higher and lower levels. To minimize confusion between the two displays, iconic vibration patterns that were designed to metaphorically represent each parameter were used to communicate their levels. The MAP levels were announced on the arm with a “heartbeat” pattern that consisted of two vibration pulses: a 300-ms pulse at a higher intensity that was mapped to systolic blood pressure, followed by a 200-ms pulse at a lower intensity mapped to diastolic pressure. The resulting sensation felt very much like the “lub-dub” characterization of heart sounds during the cardiac cycle. Figure 5.5 shows the metaphorically-derived pattern used to relate ETCO2 levels. This pattern modeled the profile of vibration intensity after the shape of the capnographic waveform during exhalation.

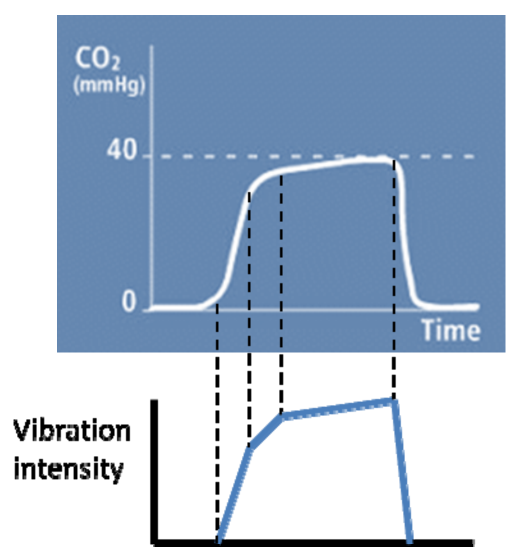


Figure 5.5: The metaphorically-derived pattern used to relate ETCO2 levels was designed to feel like the capnographic waveform.

Minimizing perceptual and cognitive interference

The minimization of perceptual interference was one of the primary motivators for employing the tactile channel for the continuous display of patient data. The clinical environment and the task set of anesthesiologists heavily load the visual and auditory channels, leaving the tactile channel less likely to face interference at the perceptual processing stage (Wickens, 2002; 2008). However, while intermodality interference can be minimized with tactification displays, the complexity of these displays makes them susceptible to perceptual interferences within the tactile channel, between display components. For example, the coincidental presentation of MAP changes on the arm and ETCO2 changes on the back could be subject to interference, causing one or both changes to be missed.

Vibrotactile adaptation and vibrotactile masking are two forms of perceptual interference that were considered in the tactification design process. Adaptation refers to the tendency for tactile sensitivity to decline with persistent vibration exposure (e.g., Pasquero, 2006; Verillo & Gescheider, 1977). Vibrotactile masking describes the decreased ability to identify a target stimulus when it is preceded or succeeded by a masking stimulus, within a short time window (e.g., Cholewiak & Craig, 1984; Craig, 1983; Pasquero, 2006). These effects are the main reason why a truly “continuous” vibration display – one that uses persistent vibration – was deemed to be impractical. Instead the tactifications were designed to present patient health data in a “semi-continous” fashion.

The best way to minimize the negative effects of vibrotactile adaptation and masking is to increase the interval between the successive stimuli. However, interstimulus intervals (ISI’s) that are too large limit the informativeness of the displays, because less information can be presented per unit time. For the respiration components, this did not present too much of a problem since the time between respirations was dictated by the respiration rate in order to best support natural mapping. The default RR setting on mechanical ventilators is (usually) 15 breaths per minute, and this is also a good average rate while patients are still awake and spontaneously breathing, therefore the respiration components of the tactification display were presented roughly every 4 seconds.

It was more difficult to determine the proper ISI for blood pressure information. Blood pressure measured via an a-line updates with every heartbeat, so new information could be available to present every second or faster. Earlier studies (e.g., Ng & Man, 2004) reported, and pilot tests for earlier versions of the tactification displays confirmed, that people generally do not like vibration presentations which occur this frequently. Therefore, with a general goal of increasing ISI, the literature was consulted to find an acceptable upper ISI limit that would not significantly degrade performance in recognizing changes between successive presentations. Studies of tactile short term memory showed that people can remember and make accurate comparisons between somewhat complex vibrotactile patterns when the timing between the pattern presentations is on the order of 5 seconds or less (e.g., Bowers, Mollenhauer, & Luxford, 1990; Gallace, Tan, Haggard, & Spence, 2008). However, some evidence suggests that this performance gets significantly worse between 2 and 5 seconds (Gallace et al., 2008: Experiment 2). With these findings serving as guidelines, and additionally considering the need to minimize overlap between the MAP and respiration display components, the decision was made for the tactifications to present MAP signals every 3 seconds. Because these signals took a little over half a second for a full presentation, this resulted in an ISI on the order of 2 seconds.

Another potential interference effect that needed to be considered is tactile change blindness, which refers to the surprising difficulty in detecting even rather large changes in a tactile pattern when the change coincides with another tactile event. This effect is similar to the masking effect, but extends beyond perception into some of the later stages of processing. To date, this effect has been demonstrated for the detection of changes along two dimensions of a vibrotactile pattern: the number of simultaneous presentation locations (e.g., Gallace, Tan, & Spence, 2006; 2007), and vibration intensity (see Chapter 2: Part A). Because vibration intensity was an important dimension in the design of the tactification displays, steps needed to be taken to minimize the negative effects of change blindness. First, since gradual intensity changes in a continuous vibration signal were found to be especially susceptible to the effects of tactile change blindness, discrete steps in intensity were employed rather than a continuously-mapped format. Secondly, the study showed that larger intensity changes (6 dB) were more resilient to these effects than smaller (3 dB) changes; therefore, the larger changes were used for these discrete steps. Finally, since presentations on the arm and the back would frequently overlap temporally, and thus be susceptible to change blindness effects, it was clear that relying solely on the dimension of intensity to relate important information would likely not be effective. Therefore, for the two components that used the intensity dimension to communicate level

(MAP and TV), redundant ways to communicate level were built into each signal so that the receiver did not have to rely entirely on detecting changes in intensity. For MAP, the dimension of spatial location redundantly communicated level. The level of TV was redundantly communicated by spatial location as well as the number of pulses in a presentation sequence and the cadence of this sequence. For each parameter, the redundancy gain reduced the likelihood of missing important changes in level.

In addition to addressing possible perceptual interference effects, it was important also to avoid cognitive interference and support efficient interpretation of signal changes. This was achieved partially through the above described natural mappings, but care was taken also to avoid interference with the cognitive processing associated with ongoing tasks. Chapter 2: Part B demonstrated how the overlap between processing codes required for interpreting tactile displays (e.g., whether spatial or nonspatial patterns were used to encode information) and those engaged for ongoing tasks can significantly impact performance on a multitask set, and thus warrant consideration in tactile display design.

The task set of an anesthesiologist is very complex. Several tasks emphasize the spatial processing code, such as manual interaction tasks that require a high degree of hand-eye coordination. Examples include laryngoscopy, installing an endotracheal tube, and delivering an intravenous drug via syringe. However, there are also a large number of tasks that engage nonspatial (i.e., categorical; Wickens, 2002; 2008) processing resources. For example, anesthesiologists are frequently required to perform mental arithmetic to calculate drug dosages or to determine the proper settings for mechanical ventilation, drug infusions, and other patient care devices. Because it is difficult to predict when the task set of an anesthesiologist requires these different types of processing, once again the best course of action is to use redundancy in encoding methods for the tactile displays. Therefore, steps were taken to ensure that the critical information in the tactification designs was encoded via both spatial and nonspatial methods for every parameter. For example, a change in MAP levels involved a change in the spatial location of the signal, but also a change in the intensity of the signal, and sometimes other aspects of the signal waveform. The expectation was that, when ongoing tasks required a high degree of spatial processing, anesthesiologist could emphasize nonspatial processing resources when interpreting MAP levels, and vice versa.

Facilitating a transition from attentional "background" to "foreground"

Qualities of the pulse oximetry display which make it ideal for physiological monitoring include its ability to exist in the attentional "background", keeping the anesthesiologist peripherally aware of the state of each parameter under normal conditions, and its ability to transition to the attentional "foreground", through a natural procession of its signal, when a serious health situation is realized. This transition usually occurs when the continuous signal presents a sufficiently large change in one of its display dimensions: tone presentation frequency (relating heart rate) or tone pitch (relating blood oxygenation). In this sense, the display enjoys all of the attention-orienting properties of an alarm display, but has the added benefit of keeping the anesthesiologist more informed during "normal" situations and during periods where the signal is "trending" toward a serious situation, while minimally adding to their mental load.

Following this model, the two tactification displays were designed to relay the state of each parameter in a relatively subtle manner when conditions were normal. This allowed anesthesiologists to be peripherally aware of the fact that the levels of important physiological measures were stable, providing a reassurance of "true negative" states (a valuable component of continuously informing displays; e.g., Sanderson, Liu, & Jenkins, 2009) without unduly capturing attention. When changes in parameter levels occurred, the two tactification displays used different methods to communicate these changes, representing different mechanisms for facilitating the transition of the displayed data from the periphery to the focus of attention. One followed the normal progression of its signal, e.g., when MAP levels decreased, the "heartbeat" signal presentation decreased in intensity and was presented from a tactor that was located lower on the arm. The other display did this as well, but differed in that it changed a part of its presentation pattern in a way that mapped the salience of the pattern to the severity of the situation. The former design is modeled more closely after the characteristics of the pulse oximetry display, while the latter represents a so-called "hybrid" display that combines the benefits of a continuously informing display (e.g., heightened awareness of the displayed parameters in normal conditions) with those of a multistage or graded alarm display (e.g., reliable reorientation of attention via exogenous mechanisms when a reorientation is warranted).

TACTILE ALARM AND TACTIFICATION DISPLAY DESIGNS

This section describes the three tactile display configurations which were designed to relate data regarding the parameters MAP, ETCO2, and TV. The first, called the “alarm” display, represented the simplest form of tactile aid, alerting anesthesiologists via a vibration presentation when one of the three parameters had surpassed high or low thresholds that defined “problem” levels. This display was evaluated alongside the tactification displays to determine the relative benefits of a simpler, but less informative tactile display. The findings of the study described in Chapter 4 suggested that monitoring performance may benefit the most from a more informative display, but performance in secondary tasks was better with a simpler display. The inclusion of the alarm display configuration in the evaluation study described later in this chapter allowed the question of the relative benefits associated with display informativeness/complexity to be revisited, in the context of a considerably more complex task environment.

The other two displays, called the “continuous” and “hybrid” displays, were tactifications. Both displays relayed MAP levels on the arm every 3 seconds. Immediately following each breath, TV levels then ETCO2 levels were relayed on the back in a way that mapped to inhalation and exhalation. Respiration rate (RR) could also be inferred by the frequency of each respiration presentation as well as the overall speed of the inhalation/exhalation pattern. In addition, the hybrid display shared some of the properties of the alarm display, notifying the receiver when parameter levels surpassed graded thresholds, and mapping the salience of the display signal to the severity of the patient’s health conditions. Figure 5.6 summarizes the main differences between all three displays.

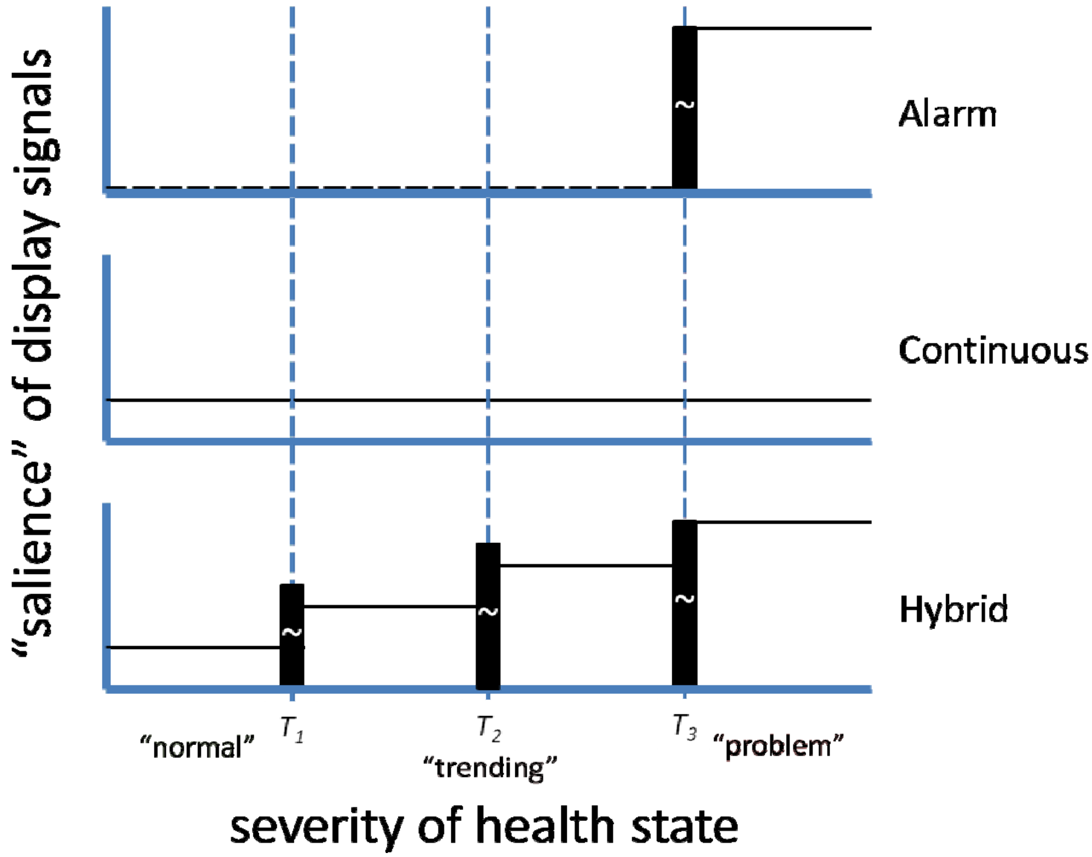


Figure 5.6: Signal adaptations with respect to changing health state. The bars labeled (~) designate "beat" pulses (see Figure 5.7) which were presented when threshold levels were surpassed.

Alarm Display

The alarm display configuration presented a signal only when either high or low threshold levels for a given parameter were surpassed. A special type of vibration pulse, referred to as a "beat" pulse, was used for alarm presentations. This pulse was created through an additive combination of two activation frequencies which resulted in an amplitude-modulated signal that gave a "rolling pulsation" sensation (see Figure 5.7). For the alarm display, the resultant frequency of these "beat" pulses was 3 Hz. The maximum intensity for these presentations was approximately 20 dB above sensory threshold, and each alarm presentation took a duration of 1000 ms.

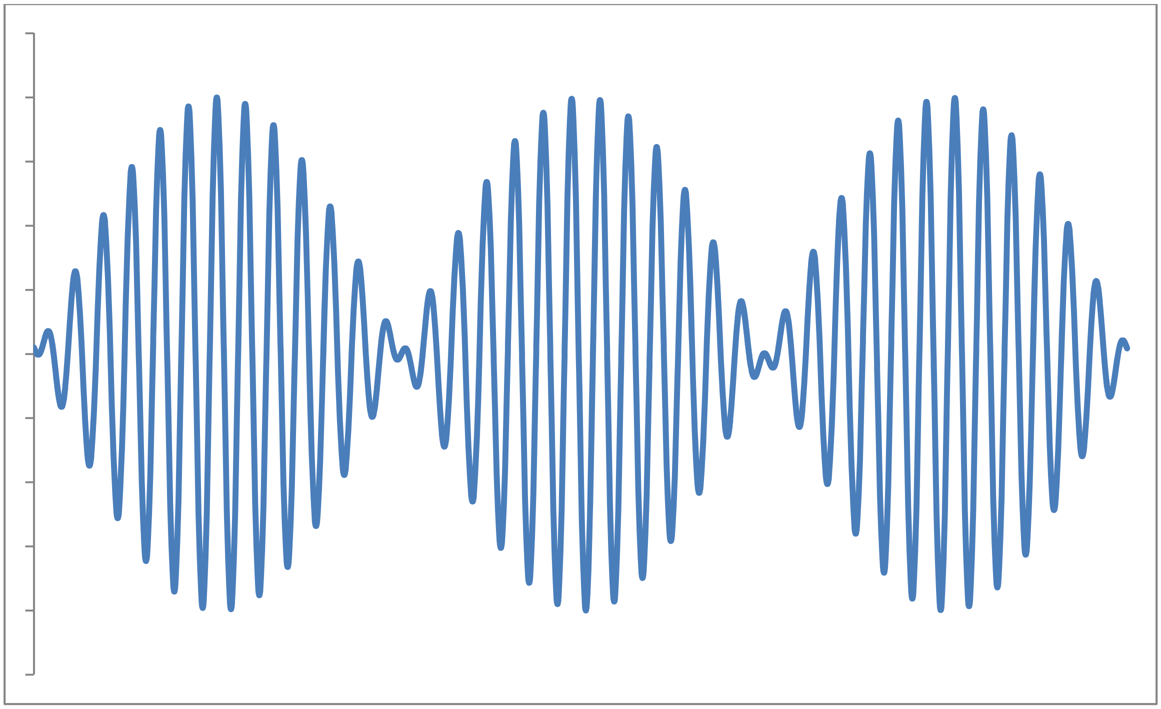


Figure 5.7: An example of the “beat pulse” signal activation pattern. The y-axis represents the amplitude of the signal.

For the MAP and ETCO2 components of the alarm display, the beat pulse was presented from the top tactor on the arm/center of the back when the upper threshold was surpassed, and the bottom tactor when the lower threshold was surpassed. The TV component presented this pulse from the bottom tactors of both “lungs” (see Figure 5.4) when TV levels were too low. When TV levels were too high, the top and bottom tactors for each lung were activated, creating the “image” of high pressures exerted over the entirety (top and bottom) of the “overinflated” lungs.

MAP alarm presentations would occur immediately when MAP levels reached “problem” levels (see Figure 5.6), then would be repeated every 20 seconds that MAP remained in a problem state. This 20 second delay for alarm repetition is similar to those of most auditory alarm systems in the clinical setting. In contrast, alarm presentations for TV and ETCO2 could occur as frequently as each breath. When both TV and ETCO2 were in a problem state, the TV alarm was presented first, followed by the ETCO2 alarm, without overlap.

Continuous Display

Each tactification display (the continuous and hybrid displays) used a similar set of rules to relate the levels of MAP and ETCO2 along their respective axes of 5 tactors. Under "normal" conditions, the vibration was presented from the tactor located at the center of the axis (i.e., the 3rd tactor from the top/bottom). The presentation of each vibrotactile icon lasted 500 ms. When parameters levels ventured outside of these normal conditions to a "trending" state (exceeding threshold T_1 in Figure 5.6), the iconic signals were instead presented from tactors immediately above or below the center tactor. When levels reached a "problem" state – the same levels that activated alarm signals in the alarm display configuration –, the presentations came from the top or bottom tactors of these axes.

Levels were redundantly communicated for the MAP parameter by modulating intensity. Under normal conditions, the maximum intensity of the icon was approximately 14 dB above the sensory threshold. Because 6 dB steps were found to be more resilient to the effects of change blindness than 3 dB steps (Chapter 2: Part A), only 6 dB changes in intensity were used. The maximum intensity that could be produced by the tactors was 20 dB, thus this level was used as the "high MAP" intensity, and 8 dB was used to communicate low MAP. The transition between intensity levels was triggered halfway through the trending phase shown in Figure 5.6, at threshold T_2. In the evaluation study, this corresponded to halfway between the boundaries of "normal fluctuation" in MAP and the "acceptable range" (see Table 5.1). Note that, since all level presentations during this trending phase came from the 2nd/4th tactor of the MAP display, the intensity transition occurred without a change in the spatial location of the signal. This improved the ability to detect the intensity transition, and also allowed the display to communicate levels with a higher resolution, as it served as an additional threshold for signal change.

For ETCO2, levels were redundantly communicated by the "steepness" of the increase in the iconic signal waveform (Figure 5.5). For example, higher levels of ETCO2 caused the pattern to increase to its maximum intensity faster than it did for lower ETCO2 levels. The maximum intensity of this vibration was always approximately 20 dB above threshold, regardless of level.

To increase the likelihood that changes in level would be recognizable for MAP and ETCO2, it helped to give a short reference signal representing where the levels were immediately before the change. This practice is used in sonification design, when longer time periods between successive signals make it difficult to compare them (e.g., Watson, 2006). Whenever a transition occurred between level thresholds (i.e., when T_1 , T_2 , or T_3 were crossed, from either direction), a 300-ms “change pulse” was presented from the spatial location associated with the previous level, immediately followed by the characteristic icon presentation which relayed the new level. For MAP presentation, the change pulse was also characterized by the intensity associated with the previous level, so that the change in intensity that occurred at threshold T_2 could more easily be recognized.

The TV presentation was designed after a metaphor of “filling” the two lungs by presenting sequential vibrations that began at the bottom tactors for each lung and continued to upward to the other tactors. Following this metaphor, the number of tactors that activated, the spatial location of the topmost vibrations, and the cadence or rhythm of the “fill” pattern redundantly communicated the level of TV. Additionally, as the fill pattern travelled upward, the vibrations representing each lung section increased in intensity, thus naturally mapping the sensation of increasing pressures on the lung walls as they inflated. Normal TV levels were represented by a pattern that filled the first 3 sections of each lung with an even cadence, i.e., the entire duration of the fill pattern (which was dictated by RR) was evenly divided between the vibrations at the 3 locations. Figure 5.8 provides an illustration of the fill pattern for normal TV levels.

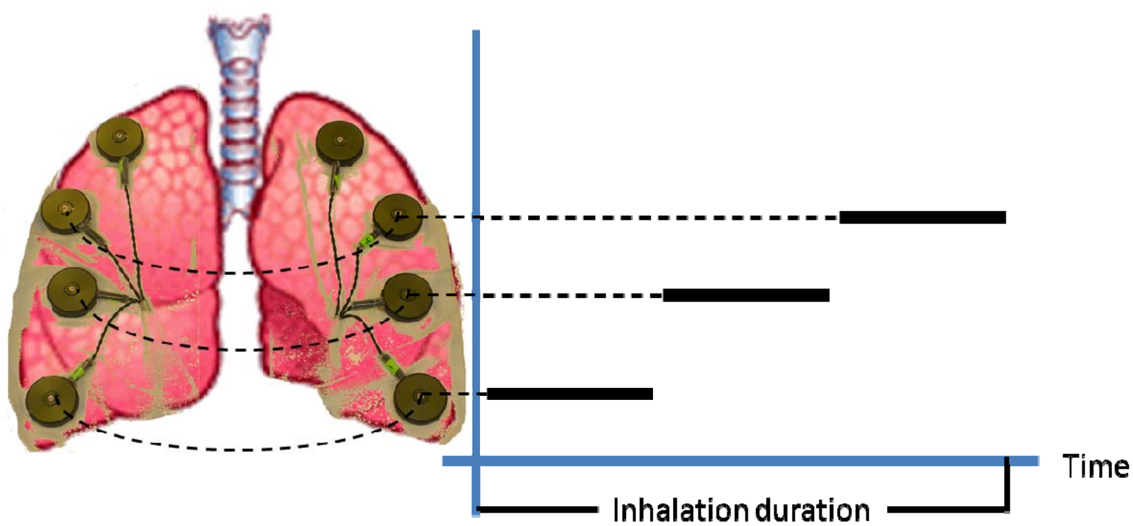


Figure 5.8: The “fill” pattern representing normal TV levels. Inhalation duration was calculated according to RR.

Following the same metaphor, TV levels that were lower than what was considered normal (i.e., crossing the T_1 threshold in Figure 5.6) took a longer time to fill up the first section of lung, and only filled to the 2nd section. Low TV levels that constituted a "problem" state (crossing the T_3 threshold) included only a single, long vibration at the bottom-most tactors. Figure 5.9 illustrates each of these presentations. TV levels that were higher than normal but still considered to be within the trending phase were related with a fill sequence that travelled more quickly through the first three lung sections and reached the 4th section of the lungs. Finally, high TV levels that constituted a "problem" state also included a fill pattern that reached the 4th section of the lungs, but arrived at this last section much more quickly, and the presentation from the highest tactors lasted twice as long. Additionally, because TV level was mapped to vibration intensity, when high TV levels constituted a problem state, the vibrations at the highest section of the longs were at the maximum intensity level of 20 dB. Figure 5.10 illustrates the fill patterns for "trending" high and "problem" high TV levels.

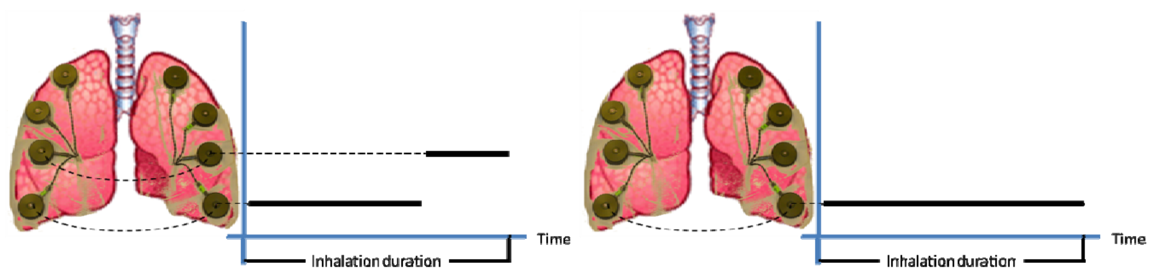


Figure 5.9: Fill patterns associated with "trending" low TV levels (left) and "problem" low TV levels (right).

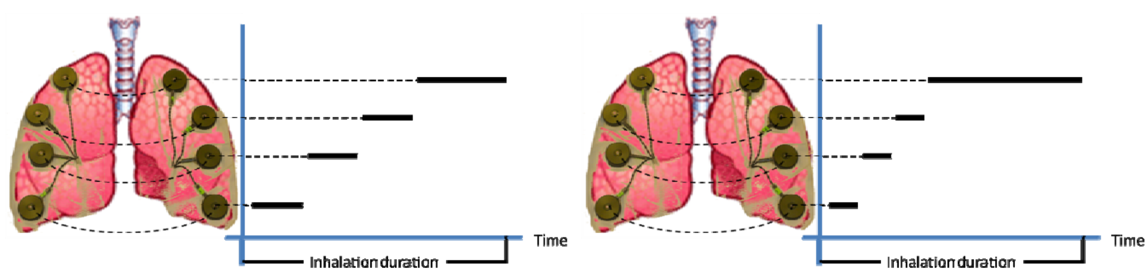


Figure 5.10: Fill patterns associated with "trending" high (left) and "problem" high (right) TV levels.

The "inhalation duration" for the TV presentation was calculated to be 3/8 of the entire interval between breaths. When RR was 15 (the default ventilator setting), this duration was

1500 ms. A 50 ms blank interval separated the vibrations associated with the fill pattern and the ETCO2 presentation. The duration of the ETCO2 pulse was at least 500 ms long, with an additional variable duration that represented the “ramp up” portion of the ETCO2 icon. This additional duration depended on RR (less for for higher RR) and the ETCO2 level (low levels made the “ramp up” take longer). Finally, a 300-ms change pulse preceded the ETCO2 presentation when appropriate. In all, the entire duration of the inhalation/exhalation pattern never exceeded 3/4 of the interval between breaths, and was under most conditions close to 1/2 of this interval.

Hybrid Display

The hybrid display followed all of the presentation rules explained for the continuous display, with the following exceptions:

- Whenever parameter levels moved away from the “normal” state and crossed the T_1, T_2, or T_3 thresholds in Figure 5.6 (i.e., when health state got worse, but not when levels crossed these thresholds while moving back toward normal), the characteristic pulse for that parameter was replaced with a “beat” pulse similar to that used in the alarm display (see Figure 5.7). For MAP and ETCO2, the beat pulse replaced the vibrotactile icon; for TV, beat pulses characterized the final vibration of the inhalation “fill” pattern.
- For the MAP display, the beat pulse signal was presented as soon as possible after a threshold had been crossed. As long as presentations did not overlap, the strict rule of maintaining a 3-second interval between presentation onsets was relaxed in these instances. For the two respiration parameters, the beat pulse signal characterized the relevant parameter(s) during the next scheduled breath presentation.
- Following a beat pulse presentation, provided another threshold wasn’t crossed before the subsequent presentation and the parameter level did not constitute a “problem” state, the subsequent presentation returned to the regular iconic pattern (i.e., heartbeat/ETCO2 waveform/inhalation fill pattern, without the beat pulse). When a

parameter level constituted a problem state, every presentation related to that parameter included the beat pulse.

- The beat pulse signal communicated an increasing urgency in the signal when crossing thresholds farther away from the "normal" state. When crossing T_1, the beat pulses occurred with a frequency of 2 Hz. Crossing T_2 was announced via signals with 3 Hz pulse frequencies, and crossing T_3 was announced with 5 Hz pulse frequencies.
- The measures of MAP and ETCO2 maintained the mapping between level and presentation location that also defined the continuous display, but abandoned the mappings between intensity/duration and level. Instead, a universal rule was followed that levels which described "worse" states involved higher intensities and longer presentations. The result of this rule was that the salience of the signal was mapped to the severity of the health state, rather than the absolute level of the parameter. For example, rather than relating low MAP levels with lower vibration intensities (as the continuous display did) the intensity *increased* (but was presented from the same tactor location, lower on the arm) to communicate a more severe health state with regard to that parameter. Similarly, the duration of the ETCO2 pulse increased from 500 ms up to 750 ms, with the increase in the duration defined by the distance ETCO2 levels were from the "normal" state.

HYPOTHESES

The following hypotheses regarding workload and performance effects guided the evaluation of the three tactile displays described later in this chapter. The hypotheses are listed in order of how they address each of the 6 goals for an improved patient monitoring display (Ch. 1). A brief explanation of the reasoning behind each hypothesis follows its statement.

1) *Improve, or at least not reduce, detection of abnormal/critical levels and dynamics in a physiological parameter*

H1: Participants will show higher detection rates for critical health events when equipped with any of the three vibrotactile displays, when compared to a baseline (i.e., visual only) display configuration.

Without the aid of tactile displays, detecting the health events (at least the ones that do not reach the threshold to trigger an auditory alarm) requires endogenously-driven overt shifts in attention and body posture to sample the visual displays. Because these shifts are detrimental to other tasks associated with patient care, participants will be motivated to minimize the number of attentional shifts, leading to a higher likelihood of missing an event. The tactile displays support awareness of physiological levels without requiring a reorientation of visual attention, thus reducing the performance costs associated with interrupting other patient care tasks. They can also serve as an exogenous attention orientation mechanism that reduces the need to rely on solely endogenous mechanisms to achieve this level of awareness.

H2: Participants will show higher event detection rates when equipped with the alarm and hybrid displays than with the continuous display.

This hypothesis reflects the expectation for a higher likelihood of missing changes with the continuous display due to tactile change blindness or perceptual masking effects. For example, changes in the blood pressure display may coincide with the dynamic presentation of a breath in the respiration display (similar to the “mudsplash” conditions in the tactile change blindness study described in Ch. 2). Because the continuous display lacks the salient “beat” pulses, which differ from the standard signal according to multiple signal dimensions, problem states are more likely to be masked due to these perceptual effects. The alarm and hybrid displays, which include these types of vibration pulses, should be more resilient to potential masking effects, therefore allowing problem states to be detected more reliably.

2) *Support earlier diagnosis of and response to developing health events, before parameters reach critical levels*

H3: Participants will detect events earlier in their development when equipped with any of the three vibrotactile displays, when compared to the baseline display configuration.

By facilitating a greater level of continuous awareness of critical physiological parameters in a way that minimally competes with ongoing tasks, participants should be aware of developing events sooner with the vibrotactile displays.

H4: Participants will detect these events earlier when equipped with the continuous and hybrid displays than with the alarm display.

This hypothesis is based on the fact that the alarm display does not have the resolution of the continuous and hybrid displays. The alarm display only changes its state in relation to two thresholds: high and low parameter levels that, when crossed, signify that a problem state has already been realized. In contrast, the continuous and hybrid displays support the early detection of trends in parameters while their levels are still within the alarm bounds. In cases where these trends are signs of, or precursors to, developing health events, a greater awareness of the trending behavior allows faster identification of problem states before or after they are realized.

H5: The performance benefits observed with the continuous and hybrid displays, in terms of earlier event detection, will be greater (i.e., earlier detection) for slowly-developing than for quickly-developing events.

Events that develop very quickly will spend relatively little time in a “trending” state before arriving at a “problem” state, while more slowly-developing events will show longer trending patterns. Detecting events in a trending state, instead of a problem state, will therefore represent a larger advantage in detection time for the slowly-developing events, compared to the advantage found for quickly-developing events.

H6: Participants will complete corrective actions in response to these events faster when equipped with any of the three vibrotactile displays, when compared to the baseline display configuration.

By improving awareness of the levels and dynamics of physiological parameters associated with health events, participants equipped with the tactile displays can track the effects of corrective actions in a way that minimally interferes with other ongoing tasks. This leads to a greater likelihood of detecting when additional actions are needed, and faster recognition of this need. For example, a drug administration meant to correct an abnormal parameter may result in an underdosing or overdosing. Participants who are more aware of the level of this parameter following the drug administration can more quickly determine when an additional correction is required.

H7: Participants will complete corrective actions faster when equipped with the continuous and hybrid displays than with the alarm display.

The continuous and hybrid displays, by virtue of their higher display resolution, will allow more precise tracking of parameter levels following corrective actions than will the alarm display. This allows for faster recognition of under- and overdosing, and any other signs of incomplete correction of a health event.

3) *Support more precise management of displayed parameters within prescribed ranges (through pharmacological or other physical interventions)*

H8: Participants will manage MAP, ETCO2, and TV levels more effectively when equipped with any of the three vibrotactile displays, when compared to the baseline display configuration.

Tactile displays of these three parameters will support earlier detection of abnormal levels in the parameters, and earlier correction of events that caused these levels. Therefore, the levels of MAP, ETCO2, and TV will spend less time at abnormal/"problem" levels, and even when these abnormal levels are realized, the parameters will reach less severe levels than they would without the aid of the tactile displays.

H9: Participants will manage these levels most effectively with the hybrid display.

This display will best support physiological management because it combines display elements that best support: baseline event detection (the salient "beat" pulses signifying critical changes in parameter levels; see H2), earlier event detection (because it supports signal gradation for detection of early trends; see H4), and faster correction of health events (because it supports more precise tracking of the effects of corrective actions; see H7).

4) Reduce competition for visual resources

H10: The patient induction task will be completed faster when participants are equipped with any of the three vibrotactile displays, when compared to the baseline display configuration.

Several of the steps required in patient induction require periods of uninterrupted visual attention. With tactile displays relating some of the critical patient information, participants will need to interrupt these tasks less often to visually sample physiological data displays. Fewer interruptions of the induction task steps will result in faster overall completion.

H11: The patient induction task will be completed fastest with the hybrid display.

This hypothesis reflects the expectation that the alarm display will lead to more unnecessary interruptions of the induction task than the hybrid display. Several "benign" events will cause tactile alarms to be activated when no real problem is present, recreating the prevalence of nuisance alarms in the clinical setting (described in Chapter 1). The greater contextual awareness supported by the hybrid display, which stems from the ability to track trends in the patient's physiology, will aid in the recognition of these nuisance alarm states and allow the participant to continue the induction task without interruption.

The hybrid display is also expected to lead to fewer interruptions of the induction task than the continuous display. This expectation is based on the fact that participants will have a harder time distinguishing problem states from normal states with the continuous display, since it lacks the distinguishing beat pulses that more clearly denote level changes by their severity. The continuous display will therefore lead to an increased likelihood of participants interrupting the induction tasks to, for example, “double-check” visual data displays to verify that an assumed change from a high level to a normal level was not instead a change from a normal level to a low level.

5) Reduce, or at least not increase, overall cognitive workload

H12: The demand imposed by scenario conditions will affect detection and correction times in a way that does not vary due to display configuration.

If a display configuration imposed a higher level of cognitive workload on participants, it could be expected that the effect of this imposed workload would be observed in appreciably worse performance responding to events during already “high demand” conditions, when compared to “low demand” conditions. It is expected that the difference in performance between high and low demand conditions will not be larger when any of the tactile displays are used than in the baseline configuration.

H13: Participants will perceive and rate mental demand and effort (measured by NASA-TLX) to be the same for all display configurations.

Because the tactile displays are not expected to impose a significant increase in cognitive workload, higher ratings for these three attributes are not expected for scenarios under any tactile display configuration.

6) *Provide better support for attention and task/interruption management by assisting anesthesiologists in making informed decisions about whether, and when, to switch attention between physiological monitoring tasks and other ongoing tasks*

H14: Scores for combined performance on both the induction task and the physiological management task will be better when participants are equipped with any of the three vibrotactile displays, compared to the baseline display configuration.

Such a "multitasking performance" score is the best quantifiable measure of the ability of a display to support attention and task management. A high score requires allocating attention efficiently: maintaining a high level of awareness of the patient's physiological state while focusing for extended periods on the visually-intensive induction task. It also requires switching between tasks only when it is necessary to do so, because each switch is costly to overall performance (e.g., Trafton & Monk, 2008). By engaging separate attentional resources, the tactile displays are expected to reduce the need to switch attention between tasks/visual displays, and thus reduce the overall performance costs. By keeping the participant more aware of physiological parameter levels, they will also support effective decision making about when a task switch – to manage physiological levels – is warranted, and when it is more important to remain focused on the induction task. These advantages will be reflected in higher overall multitasking performance scores under tactile display configurations.

H15: Combined performance scores will be best with the hybrid display configuration.

The hybrid display is expected to offer the greatest advantage to multitasking performance because, of the three tactile displays, it should best support both the physiological management task (see H9) and the induction task (see H11). Attention should be most effectively managed with this display because it offers the opportunity to endogenously orient attention to sample critical physiological parameter levels in a manner that minimally competes for the attentional resources required by other tasks. Additionally, it supports exogenous attention orientation when parameter levels/dynamics are severe enough to warrant capturing attention through the beat pulse characteristic. This characteristic also makes the severity of the situation easier for the participants to assess before making decisions about whether or not to switch tasks, thus most effectively supporting task management.

H16: Participants will subjectively rank the baseline display configuration as worst, and the hybrid display configuration as best, in terms of the ability to support multitask performance.

Participants are expected to recognize the performance benefits of the tactile displays, and since the hybrid display is expected to lead to the best multitasking performance, it is expected that participants will rank it as the highest for this attribute.

H17: Participants' subjective rankings of the alarm configuration's ability to minimize unnecessary distraction will be worse than for the continuous or hybrid display configurations.

Because the alarm display does not support the level of continuous contextual awareness that the continuous and hybrid displays do, it is expected that benign events which trigger nuisance alarms will lead to a higher likelihood of unnecessarily distracting participants from the induction task under the alarm configuration.

EVALUATION STUDY

A common limitation of studies evaluating advanced physiological display technologies is that these evaluations do not involve a task set that imposes any significant workload by secondary tasks (Görges & Staggers, 2007; Sanderson, Watson, & Russell, 2005). Even when secondary tasks are present, they often only remotely mirror clinical practice. Therefore, in designing the study to evaluate the tactile displays described in this chapter, specific efforts were taken to create a task set that a) imposed a significant secondary task load to be completed alongside physiological monitoring tasks and b) accurately reflected the types of tasks that an anesthesiologist would conduct in the clinical setting. This study was set in a realistic simulation environment, used practicing anesthesiologists as participants, and required the completion of tasks that were modeled after the induction phase of anesthesia: the phase that is both the most critical for patient care (i.e., the most can go wrong) but also that which imposes the greatest challenge to effective physiological monitoring (Gaba & Lee, 1990; Loeb, 1993; 1994).

TAPIS: THInC Lab Anesthesia Patient Induction Simulator

The task set of an anesthesiologist during the induction phase of general anesthesia involves preparing the patient for surgery by administering drugs to render the patient unconscious, manage their pain levels, and paralyze the skeletal muscles. After the patient is paralyzed, the anesthesiologist needs to install a rigid tube in the patient's trachea – an endotracheal tube – which is then connected to a mechanical ventilator so that respiration can continue while the patient's diaphragm muscles are paralyzed. The anesthesiologist also attaches many of the sensors required to monitor the patient's physiology throughout the surgical procedure during this phase. Throughout all of these activities, the anesthesiologist is responsible for making sure the patient is physiologically stable. This responsibility can be especially challenging because of the unexpected ways patients can respond to the administered drugs, and also because of the need for the anesthesiologists to divide their attention between various tasks and physiological data displays. For the purpose of this study, a simulation environment was developed that recreates these attentional and task demands. Figure 5.11 shows an overhead view of the layout for the THInC Lab Anesthesia Patient Induction Simulator (TAPIS).

TAPIS was largely modeled after the layout for displays and controls/tools available to anesthesiologists in the perioperative space in the University of Michigan (UM) hospitals, with one important difference: the physiological data displays ("physiological data 1" and "physiological data 2" in Figure 5.11) were arranged to be outside of the participants' field of view whenever they were directly interacting with the patient (on the "patient display" in Figure 5.11). This arrangement enforced a requirement for an overt shift in the focus of visual attention (and posture) away from the patient any time participants wished to visually sample the physiological data. This method of prohibiting both the patient and the physiological displays from being visible within the same field of view may seem somewhat artificial. However, there are a wide range of situations which occur during even the most routine clinical cases for which it is not possible for the anesthesiologist to see physiological data displays while interacting with the patient (e.g., when tasks require going under surgical drapes or underneath the operating table). Also there are several situations where a shift in visual attention away from the patient could be very detrimental to a patient interaction task (e.g., during endotracheal intubation or when placing intravenous lines). Finally, this arrangement was justified to offset the commonly observed "simulator effect", i.e., the hypervigilance participants tend to exhibit when they know their ability to detect critical events is being evaluated.

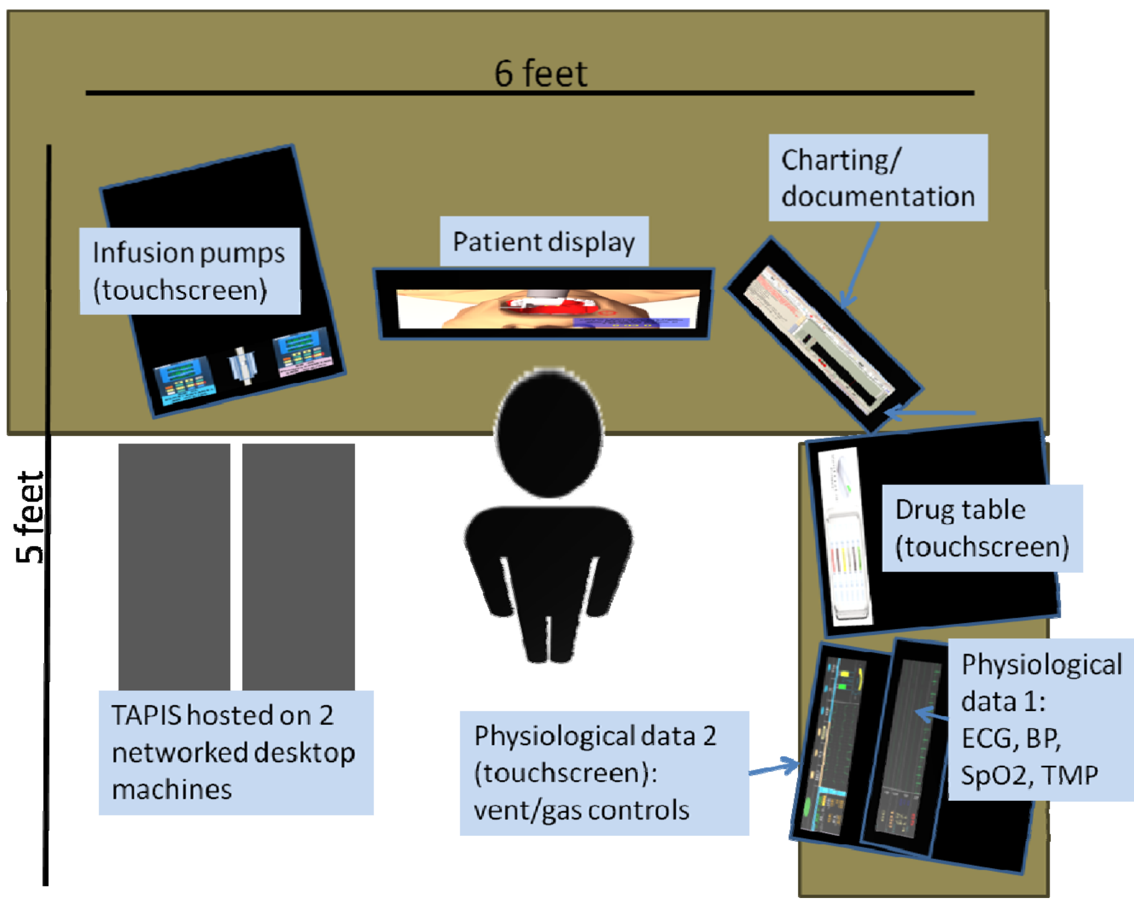


Figure 5.11: Layout of visual displays for the TAPIS simulator.

In addition to the visual displays in TAPIS, the simulator included an auditory pulse oximetry sonification which participants were expected to monitor. Also, auditory alarms associated with each physiological parameter would sound when the parameter levels surpassed the default alarm thresholds used in UM hospitals. Also modeled after the UM hospitals displays, the lower and upper alarm thresholds were visually displayed next to some of the relevant parameters (see Figure 5.17). Finally, a constant loop of background noise, recorded in a hospital operating room during a routine case, played throughout the simulation. The resultant noise level was sufficient to mask the noises caused by tactor activation.

Figure 5.12 shows the displays/interfaces within the participant's field of view when interacting with the simulated patient. Controllers for interaction with the patient display include a desktop mouse and a gyroscopic Bluetooth controller originally designed for the Nintendo Wii™, referred to as the "wiimote". The wiimote was tethered to the base of the patient display, and was used as the controller for any tasks that required being in physical contact with the patient (e.g., holding a gas mask on the patient, holding the endotracheal tube in place, controlling any tool – such as the laryngoscope – that is inserted into the patient's mouth). Additionally, a "Big Red Button" (BRB), which could easily be depressed with either hand or with the back of the hand or arm while holding the wiimote, was used to record participants' detection of health events and also to access a "Troubleshooting menu" (described later in this section). The infusion pumps display was a touchscreen interface, allowing the infusion pumps to be programmed, and the IV drip rate to be set (by moving the roller clamp position) in a natural manner. The charting/documentation display was hosted on a laptop and reproduced the documentation interface used at the UM hospitals. Interaction with this display was done via the laptop's touchpad and keyboard.

Figure 5.12: Participant field of view when facing the patient display (center). To the left are infusion pumps and drip rate controls (touchscreen), and to the right is the charting/documentation interface. The wiimote and mouse were the primary controllers for interactions with the patient.

Figure 5.13 shows the displays/interfaces within the participants' field of view when turning to the right of the patient display. Participants could easily view and interact with the "drug table" display (a touchscreen, facilitating easy selection of drug syringes and turning on the mechanical ventilator) while holding the wiimote/mouse controls used for patient interaction (Figure 5.12). This allowed, for example, drugs to be selected with the right hand while the left hand held the wiimote upright to keep the gas mask in place on the simulated patient. In contrast, viewing and interacting with the physiological data displays to the far right required physically turning around and abandoning the patient display and associated controls. To enforce this behavior, participants were instructed to avoid stretching to reach the vent controls while holding the wiimote; instead, they were told to place the wiimote down on the table anytime they wished to view or interact with the physiological data displays. The "physiological data 2" interface was designed to be identical to the ventilation display used in UM hospitals, and its touchscreen capability allowed ventilator settings to be changed in a very natural manner.

Figure 5.13: View to the right of the participant (when turned away from the patient display). Left to right are the charting/documentation interface, the drug table (touchscreen) display/interface, physiological data 1 (top) and physiological data 2 (bottom, touchscreen).

Induction task

The multitask set in this study can be described as two macro-level tasks: the patient induction task and the physiological monitoring task. The induction task required completing all of the steps in a standard Rapid Sequence Induction (RSI). RSI is a type of induction procedure which involves special steps to get the patient anesthetized and physiologically stabilized as quickly as possible. RSI was recommended for the simulated patient because of a strong tendency for gastric reflux (see the patient chart in Appendix 2); this procedure minimizes the chance of gastric contents leaking out of the stomach and into the airway when the patient is paralyzed.

Most of the interaction associated with the induction task involved the patient display (Figure 5.12, center). Step-by-step instructions were listed in the bottom corners of the screen, and color coding was used to identify tools and anatomical structures relevant to each instruction (e.g., items highlighted in yellow showed where to click with the mouse). Any interactions which required being in physical contact with the patient required button presses or changes in the orientation of the wiimote, while the mouse was used for more general activities, such as picking up and using tools common to an RSI procedure. Figure 5.14 shows views of the two tool tables. Each table was accessed by "looking" to the left or right – moving the mouse in the appropriate direction – when a tool needed to be selected or replaced.

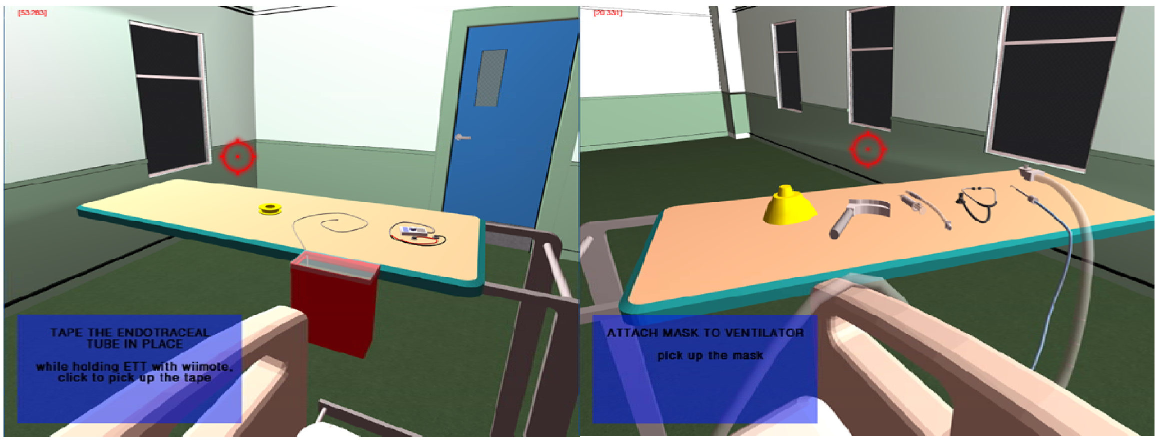


Figure 5.14: Views of the two tool tables, accessed by "turning" to the left or right of the patient, by way of moving the mouse.

The mouse was also used when two hands were required to complete an induction step. There were several such steps, for example, holding the mask over the patient's mouth (with the wiimote) while inserting a drug syringe into the IV port, depressing the syringe plunger, removing the syringe and disposing it by throwing it into the "sharps" bucket (all of which was performed with the mouse). These steps frequently required very precise mouse control and extended periods of focused visual attention. The posture that such steps imposed made it impossible to view the visual physiological data displays (see Figure 5.15), thus creating several opportunities for health events to occur which could go (at least temporarily) unnoticed without the aid of tactile displays. For some induction task steps, interrupting the step to visually sample the physiological data displays would lead to performance costs in the induction task. For example, such an interruption would require participants to release the wiimote and mouse controllers, causing the step to be restarted upon returning to the induction task. This would extend the overall time required to complete the induction task, thus participants were instructed to only interrupt such steps when they felt relatively certain that an interruption was warranted.

Figure 5.15: Induction task steps depicting the insertion of the endotracheal tube, which involved using the tethered wiimote to control the laryngoscope and the mouse to insert the tube. The requirement for prolonged visual attention during this step made it difficult/impossible to view the physiological data displays (top right corner).

In addition to the patient display, the induction task also required interacting with each of the other displays/interfaces in the TAPIS environment, to a lesser extent. For example, steps such as starting the IV drip and setting the initial infusion pump programs were completed by interacting with the infusion pumps display, adjusting gas flow settings required interacting with physiological data display 2, selecting drug syringes and turning on the ventilator required interacting with the drug table display, and documenting the steps in the RSI was done on the charting/documentation display. For every such step, the patient display would list the proper instructions and direct the participant to the appropriate display.

The complete list of all the steps involved in the RSI (constituting the entirety of the induction task) can be found in Appendix 3. Participants were told that one of their two primary goals should be to complete the induction task as quickly as possible (the other, of equal priority, was to manage the patient’s physiology, as described in the next section), and that the time it took them to complete the task was one of the primary measures of performance. Completion of all the steps in the induction task took roughly between 10 and 15 minutes, dependent largely on the amount of time participants spent on actions related to physiological monitoring and management, and how often some steps in the induction task were restarted due to interruption.

Physiological monitoring task

While concurrently performing the induction task, participants were responsible for monitoring the physiology of the patient, and using drug interventions and physical interactions to manage the patient's health. Participants were told that because the patient was undergoing a major vascular surgery called a carotid endarterectomy (see the patient chart in Appendix 2), the physiological parameters of MAP (visually displayed on physiological display 1, see Figure 5.16) and ETCO2 (see Figure 5.17) were especially critical to the patient's overall health and therefore were to be closely monitored. Additionally, since the patient was a longtime smoker and could potentially suffer from lung disease, they were also to keep a tight watch on TV levels (Figure 5.17).

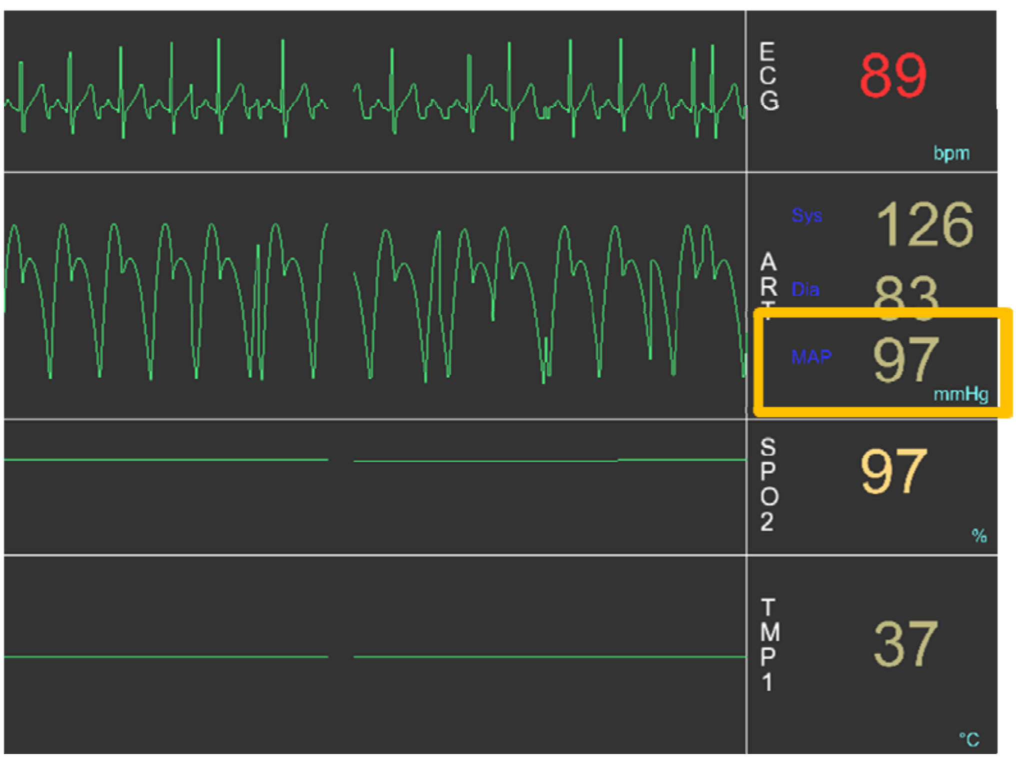


Figure 5.16: Physiological display 1, highlighting the parameter MAP, which was one of the three critical parameters for the physiological monitoring task.

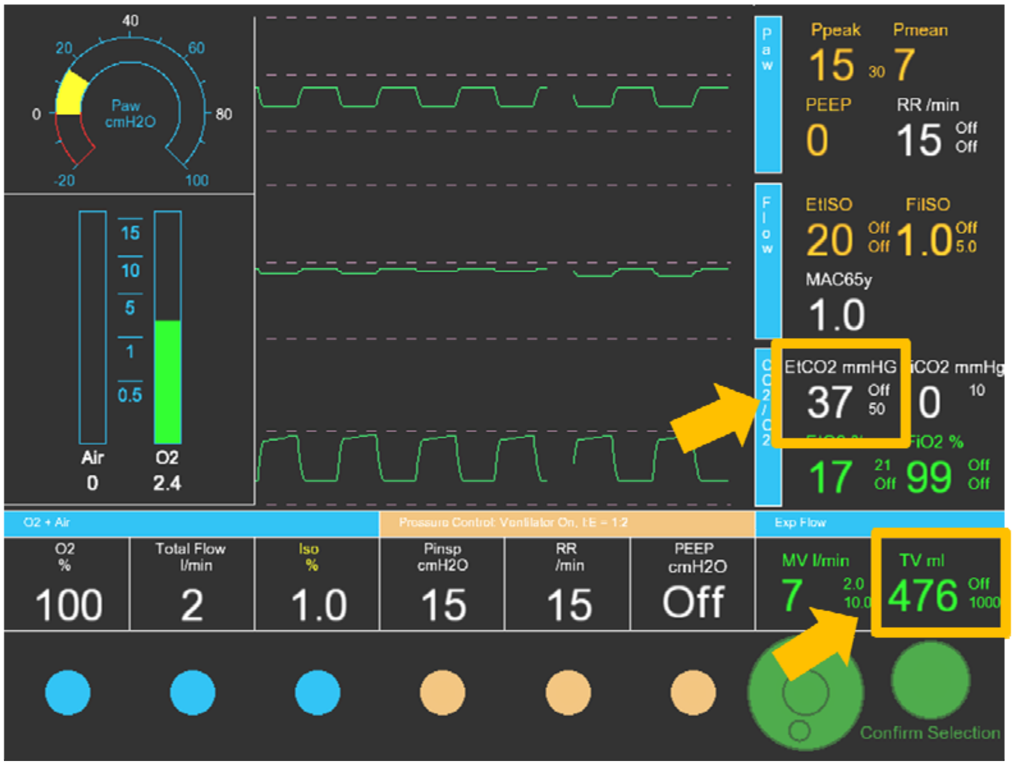


Figure 5.17: Physiological display 2, highlighting the parameters ETCO2 and TV, two of the three critical parameters for the physiological monitoring task.

The levels of each displayed physiological measure (not just the three which were most critical to the physiological monitoring task) were influenced by three factors: “normal fluctuation” scripts, “event” scripts, and simulation controls. A “normal fluctuation” script dictated baseline levels (equal to the target values listed in Table 5.1) and small second-by-second changes in each parameter for the duration of the simulation. Four versions of these normal fluctuation scripts were loosely modeled after physiological datasets that were collected and shared by colleagues and are described in Liu, Görges, & Jenkins (2010). Appendix 4 includes the normal fluctuation script from Scenario 1 as an example. Every second, the simulation would read in the next line of values corresponding to relative changes in each parameter and update the parameter levels accordingly. The changes in parameter levels dictated by these scripts were pseudo-random but controlled so that the parameters never exceeded a normal fluctuation range when only these scripts were active.

“Event” scripts added layers of changes to each parameter on top of the normal fluctuation changes. These scripts would become activated when they were triggered by related steps in the induction task. For example, the administration of each drug activated a script that dictated changes in physiological parameter levels that were representative of the drug's normal pharmacodynamic effect, and the laryngoscopy step (inserting a tool down the patient's throat to displace the tongue and visualize the airway) activated a script that caused a characteristic rise in blood pressure and heart rate. Each second, parameter levels would change by an amount that was equal to the sum total of changes dictated by all active scripts.

Some event scripts, such as the drug administration and laryngoscopy scripts, could be called benign in that they caused only slight changes in parameter levels and/or changes did not constitute a serious health event. Most of these benign scripts would play out, then after a period of time, reverse their effects so that the resultant impact on physiological levels was negligible. Other event scripts represented serious health events. These tended to cause one or more of the three critical health parameters (MAP, ETCO2, and/or TV) to reach levels outside of their defined "acceptable" ranges, and to stay outside of these ranges until participants took the proper corrective actions to return the levels back within the acceptable range. Participants were specifically instructed that their performance on the physiological monitoring task was defined by their ability to keep each parameter within the acceptable range as often as possible, and that this goal should be given equivalent priority to the goal of completing the induction task as quickly as possible. The acceptable ranges for each parameter were clearly listed and highlighted on the patient chart (Appendix 2) so participants could review them at any time. Table 5.1 lists the target levels, normal fluctuation ranges, and acceptable ranges for each of the three critical physiological parameters.

Physiological parameter	**Target value (normal fluctuation)**	**"Acceptable" range**
MAP	100 (+/-5) mmHg	85 – 115 mmHg
ETCO2	38 (+/-2) mmHg	32 – 44 mmHg
TV	500 (+25/-50) mL	375 – 600 mL

Table 5.1: Target values and "acceptable" ranges for each of the three critical parameters in the physiological monitoring task.

Note that the limits of acceptable ranges for each physiological measure were handled as the thresholds for "problem" states (the T_3 threshold in Figure 5.6). The range of "normal fluctuation" categorizes the "normal state" in Figure 5.6, with the T_1 threshold (Figure 5.6) representing the first integer value outside of this range. The T_2 threshold was set at values exactly halfway between the boundaries of the normal fluctuation range and the acceptable range.

The limits for the acceptable range were narrower than the threshold limits for auditory alarms, thus participants could not necessarily rely on auditory alarms to alert them when these ranges were exceeded. Some but not all of the serious health events eventually could lead to parameter levels that triggered these auditory alarms if left uncorrected. Several benign events also caused these alarms to sound. For example, the pulse oximetry sensor "became askew" on the simulated patient's finger at least once during each scenario, causing a "sensor fail" alarm to sound. Some of these benign events were modeled after the "nuisance" alarms that plague the clinical setting. For example, after the simulated patient had been paralyzed and apneic (i.e., not breathing) for some time, any respiration activity – which could be triggered by manual ventilation (squeezing a bag to force air into the patient) or turning on the mechanical ventilator – showed very high levels of ETCO2 for the first few breaths before settling to a normal level. Frequently, these ETCO2 levels activated the auditory alarm, though they were not characteristic of an unexpected or problematic situation.

Finally, in addition to the combined effects of all active scripts, the levels for each of the three critical parameters could be affected by the participant through simulation controls. For MAP, this was done primarily by adjusting the programmed rates of drug infusions. The drugs remifentanil and phenylephrine were loaded into the left and right infusion pumps and were to be used to lower and raise blood pressure, respectively. The concentrations of each drug were chosen so that easily programmed rates, such as those in increments of 10 mL/hr, corresponded to parameter level changes of magnitudes that facilitated easy mental calculations. To aid the participants, the effects of increasing the infusion rate 10 mL/hr on blood pressure and heart rate levels for each drug were written on a sticky note and placed below each pump (see Figure 5.18). When a new rate was programmed into the infusion pump and the "Start" button was pressed, the relative change in the rate was used to calculate the corresponding change in each parameter level. These changes took effect gradually, such that the full effect was realized after a period of 20 seconds. The pumps could easily be "paused" – which served the same effect as reprogramming a rate of 0 mL/hr – and restarted at the programmed rate with single button presses.

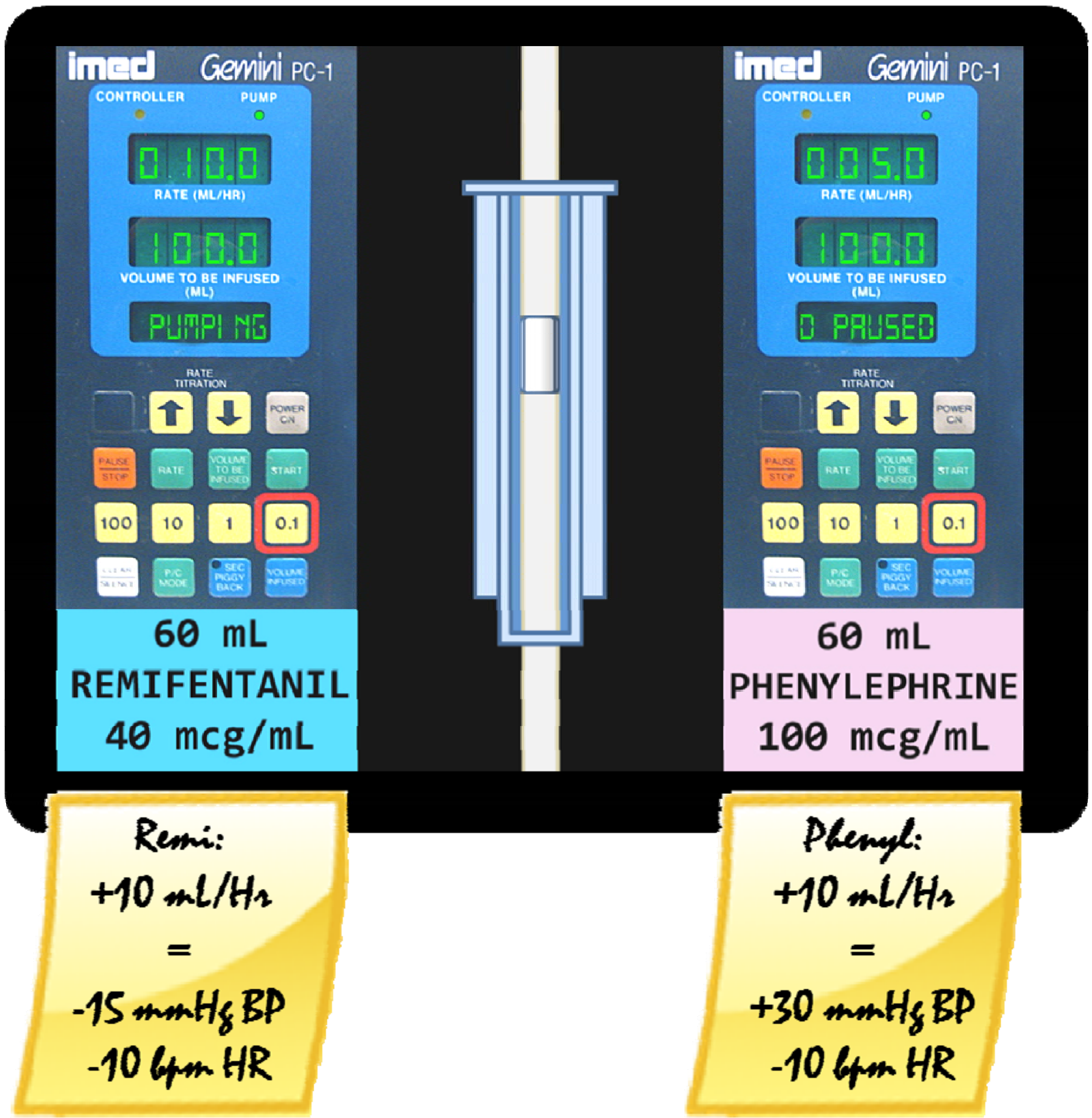


Figure 5.18: Infusion pump display (touchscreen interface). Participants programmed infusion rates of remifentanil and phenylephrine to manage MAP. Sticky notes were added below each pump to aid in mental calculations of the effects of each infusion rate on parameter levels.

When the patient's endotracheal tube was connected to the mechanical ventilator and the ventilator was turned on, the ETCO2 and TV parameter levels were affected by the ventilator settings. The ventilator was set in a standard mode known as Pressure-Controlled Ventilation (PCV), which meant that for each breath it would force air into the patient's lungs at a constant inspiratory pressure. The default inspiratory pressure was set at 15 cmH2O (see the setting for Pinsp near the bottom-center of Figure 5.17), which under normal conditions delivered TV levels right around the target value of 500 mL. Changing the Pinsp value, which could be done by participants via controls modeled after the UM hospitals system, caused TV levels to change proportionally, with higher pressures delivering higher TV levels. The change in TV corresponding to each change of the Pinsp setting could be observed in the breath immediately following the activation of the new setting. Changing Pinsp also had an indirect effect on ETCO2 levels: when TV levels changed (due to Pinsp changes or the effects of event scripts), the partial pressure of CO2 exhaled with each breath changed proportionately. This change in ETCO2 was realized over a period of several breaths, between 10 and 20 seconds. The other, more direct way to affect ETCO2 levels was to change the respiration delivery rate (RR, immediately to the right of the Pinsp setting in Figure 5.17). Under normal conditions (and with TV at normal levels), the default RR setting of 15 respirations/minute keeps ETCO2 levels right around the target value of 38 mmHg. Increasing the RR setting causes this level to drop, and decreasing RR causes it to rise, in an inversely proportional relationship. The full effect of changes in RR on ETCO2 could be realized within two breaths.

Roughly half of the serious health events could be corrected with straightforward adjustments of infusion pump or ventilator settings. The other half of these events required physical interactions with the patient or equipment to resolve the problems. These interactions were listed on a “Troubleshooting menu,” which was easily accessed at any time by pressing the Big Red Button (BRB) located in front of the patient display near the wiimote (see Figures 6.12 and 6.15). This menu included three categories of troubleshooting activities (see Figure 5.19): 1) Emergency activities such as administering an emergency drug or performing CPR; 2) Patient Interaction activities such as checking for a pulse or auscultating the lungs (listening to breath sounds with a stethoscope); and 3) Equipment Interaction activities such as inspecting the pulse oximetry sensor (which is clipped on the patient’s finger but frequently becomes askew, causing the auditory pulse oximetry signal to go silent until it is fixed), passing a suction tool into the endotracheal tube, or adjusting the position of the tube in the patient’s airway.

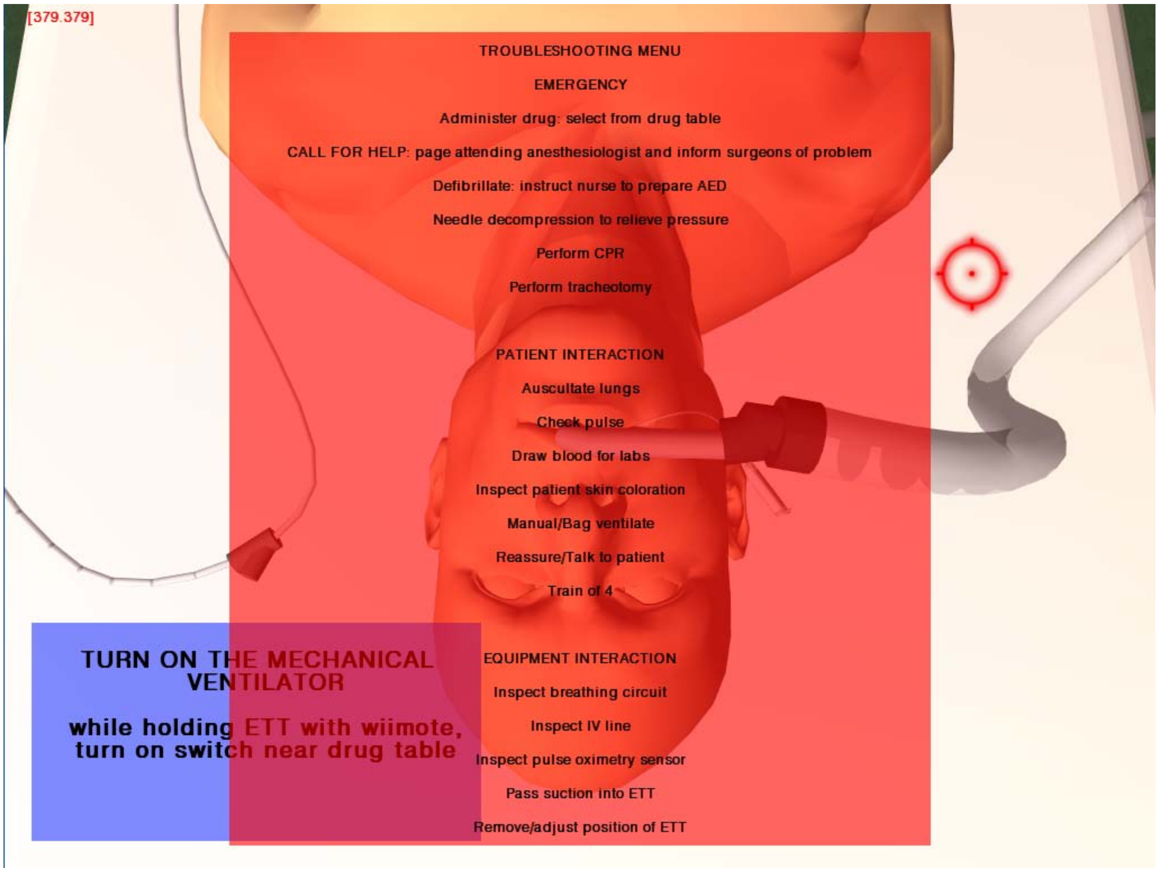


Figure 5.19: The Troubleshooting menu, which was displayed whenever the BRB was pressed. Participants could select a troubleshooting activity with the mouse or press the BRB a second time to close the menu.

Participants were told that their performance on the physiological monitoring task was defined according to: a) how quickly they detected the presence (or development of) a serious health event; b) how quickly they resolved the event; and as an extension of these two measures, c) how well they were able to minimize, for the duration of the scenario, the amount of time physiological parameter levels were outside of the acceptable ranges listed in Table 5.1 and the magnitude of the exceedance of acceptable range boundaries. Instructions were for participants to press the BRB as soon as they suspected the occurrence of a serious health event, then to verbalize the symptoms of the event so that the experimenter (present in the experimental room and playing the role of attending anesthesiologist) could hear. This could be done when parameter levels were outside of acceptable ranges, but participants were also instructed that if they detected trends in the physiological data that might eventually lead to levels outside these ranges, they may press the BRB to record the early recognition of these trends.

Participants were encouraged to press the BRB to record event detection even in cases when the event did not require a troubleshooting menu action (i.e., instead an adjustment of drug infusions or ventilator settings could resolve the problem), because this was often the fastest method to record detection time in the experimental data. In these cases, pressing the BRB a second time would cause the troubleshooting menu to disappear when participants were ready to resume the induction task. Detection time could also be calculated by noting the time until the first press of a control button on either the infusion pumps or ventilator controls, but the rule participants tended to follow was “press the BRB whenever any potential problem is noticed.”

After detecting a potential serious health event, participants were allowed to consult the attending anesthesiologist – the experimenter – for aid in resolving the event. The experimenter would then give verbal instructions regarding corrective actions for the event as quickly as possible. However, participants were made aware of the strict requirement that *all* relevant symptoms of the event (at least those related to the three critical physiological parameters) needed to be verbally reported by the participant before the experimenter would reveal the proper corrective action(s), i.e., those that would completely resolve the event. The experimenter would not tell participants when they missed a symptom; instead he would give instructions for correcting only the reported symptoms.

Simulation scenarios

Four different scenarios were created within the simulation environment. These scenarios were completed by participants in sequential order under each of the four display configurations in a randomized order. Each scenario was designed to be of equivalent difficulty, in terms of balanced occurrence of benign and serious health events. Appendix 3 includes all of the steps in the induction task, with indications of when each health event (benign and serious) was triggered during each of the four scenarios.

Serious Health Events

Serious health events were classified by the affected physiological parameters. Blood Pressure (BP) events described those that only involved changes in MAP levels. Events that involved changes in TV, ETCO2, RR, or any combination of those three parameters (but left MAP stable) were classified as Respiration events. Emergency events were those which involved abnormal levels or dynamics in MAP and at least one of TV, ETCO2, and RR. Note that while respiration rate (RR) was not one of the three critical parameters to be managed for the physiological monitoring task, some events included unusual respiration rates and rate patterns, so this parameter could be referenced in diagnosing these events. Each scenario included 2 blood pressure events, 2 respiration events, and 1 emergency event. Example scripts for each of these event types can be found in Appendix 5. Table 5.2 lists each serious health event, in order of their activation within the induction task. Also listed in this table are the event type, the scenario during which it occurred, a brief description of the event, and the proper correction to resolve the event. Appendix 3 can also be consulted to see the specific steps during the induction task in which the scripts for these events were activated.

Health event (type, scenario)	Event description	Proper corrective action
hyper_nervous (BP, 3)	Nervous, awake patient becomes hypertensive (i.e., MAP rapidly climbs) during preoxygenation steps.	Adjust remifentanil and phenylephrine infusions.
shortshallow1 (Respiration, 4)	Patient hyperventilates: each breath very short (high RR) and shallow (low TV) at beginning of preoxygenation steps. If this behavior is not corrected, it eventually leads to low ETCO2 levels.	Select “Reassure/talk to patient” from troubleshooting menu.
shortshallow2 (Respiration, 3)	Patient initially breathes normally but begins hyperventilating in a more gradual manner near the end of the preoxygenation steps.	Select “Reassure/talk to patient” from troubleshooting menu.
hypo_propofol (BP, 1)	Following administration of propofol, patient has a strong hypotensive reaction (MAP gradually falls to abnormally low levels).	Adjust infusions.

Table 5.2: Serious health events, listed in order of their occurrence within the induction task, summary descriptions of the events, and relative properties (continued on next three pages).

Health event (type, scenario)	Event description	Proper corrective action
PEA (Emergency, 4)	Pulseless Electrical Activity cardiac arrest: blood pressure trends dramatically, then goes to zero, patient stops spontaneously breathing (TV and ETCO2 levels go to zero). Pulse oximeter alarms sound as if there's a sensor failure. Choosing "check pulse" from the troubleshooting menu returns: "No pulse found!"	Select "Perform CPR" from troubleshooting menu.
hypovolemia1 (BP, 3)	Patient slowly becomes hypotensive shortly after being rendered unconscious, due to low fluid volume (i.e., patient is dehydrated due to low fluid drip rate and/or reaction to drugs).	Adjust infusions and/or increase saline drip rate.
hyper_scope1 (BP, 4)	During laryngoscopy, patient reacts strongly to stimulation of laryngoscope, rapidly becomes hypertensive.	Adjust infusions.
hyper_vomit (BP, 2)	Following insertion of the laryngoscope, gastric reflux fluid leaks into the patient's airway. This causes the patient to become hypertensive while anesthesiologist applies suction to remove the fluid.	Adjust infusions.
hyper_scope2 (BP, 1)	Patient rapidly becomes abnormally hypertensive in response to the insertion of the endotracheal tube into the trachea.	Adjust infusions.
hypovolemia2 (BP, 2)	Patient slowly becomes hypotensive after the intubation steps are completed, a delayed reaction to intravenous induction drugs.	Adjust infusions and/or increase saline drip rate.
inadequate_depth (emergency, 3)	The anesthetic drugs wear off more quickly than expected (or the patient was initially underdosed), leading to the patient showing signs of an inadequate depth of anesthesia. Immediately after the ventilator hose is attached to the intubated patient, but before the ventilator is turned on, the patient shows hypertension and signs of respiratory activity (before the mechanical ventilator is turned on, there should be no activity). The patient can be observed "coughing against the ventilator" when it is turned on, showing erratic TV and "spikes" in MAP. The patient can also be observed mildly "bucking": subtle upper-body movements with each cough.	Select "Administer emergency drug" from troubleshooting menu, select an additional dose of propofol from drug table, complete the subroutine to administer the drug.

Table 5.2 (continued): Serious health events, listed in order of their occurrence within the induction task, summary descriptions of the events, and relative properties.

Health event (type, scenario)	Event description	Proper corrective action
RMI1 (Respiration, 2)	Right-Mainstem Intubation: the endotracheal tube is initially inserted too far into the airway so that it rests in the right bronchial mainstem, rather than the trachea. This leads to inflation of only the right lung, observable as TV being roughly half of normal levels. This is first recognizable when the ventilator is turned on. Choosing "Auscultate lungs" from the troubleshooting menu would cause breath sounds to be heard only on the right side of the chest.	Select "Adjust position of endotracheal tube" from troubleshooting menu, complete the subroutine for this adjustment.
hyperventilation (Respiration, 1)	When mechanical ventilator is turned on, the default pressure settings cause the lungs to overinflate, showing high TV levels and eventually leading to low ETCO2 levels.	Adjust ventilator settings (Pinsp and RR).
hypo_isoflurane (BP, 4)	Patient has an abnormally strong hypotensive reaction to the inhaled anesthetic agent isoflurane, which is administered when the ventilator is turned on.	Adjust infusions.
hypercarbia (Respiration, 2)	Shortly after ventilator is turned on, high levels of ETCO2 result when default respiration rate settings on mechanical ventilator are insufficient to maintain normal levels.	Adjust ventilator settings (RR).
pneumothorax (Emergency, 1)	Tension pneumothorax: during train-of-4 steps, one of the lungs overinflates (very high TV levels) then ruptures, resulting in a pocket of air inside the chest cavity which puts pressure on the heart and lungs. This causes precipitous drops in TV and MAP levels until they are relatively stable but at very low levels.	Select "Perform needle decompression to relieve pressure" from troubleshooting menu.
anaphylaxis (Emergency, 2)	Patient experiences a severe allergic reaction following the administration of the muscle paralytic vecuronium near the end of the induction. The reaction develops over a few minutes, leading to severe hypotension, bronchospasm (low TV levels due to increased lung resistance), high heart rate, and fever. The capnographic (ETCO2) waveform shows a characteristic bronchospasm pattern. In later stages, a rash can be observed on the patient's chest and neck area. "Wheezelike" breath sounds are heard during the auscultation subroutine if "Auscultate lungs" is chosen from the troubleshooting menu.	Select "Administer emergency drug" from troubleshooting menu, select epinephrine from drug table, complete the subroutine to administer the drug.

Table 5.2 (continued): Serious health events, listed in order of their occurrence within the induction task, summary descriptions of the events, and relative properties.

Health event (type, scenario)	Event description	Proper corrective action
RMI2 (Respiration, 3)	Endotracheal tube slips from the trachea into the right bronchial mainstem when the esophageal temperature probe is inserted near the end of induction. This causes TV levels to suddenly drop to about half of their normal levels. Choosing "Auscultate lungs" from the troubleshooting menu would cause breath sounds to be heard only on the right side of the chest.	Select "Adjust position of endotracheal tube" from troubleshooting menu, complete the subroutine for this adjustment.
hypoventilation (Respiration, 4)	Default vent settings are insufficient and/or reaction to drugs leads to a gradual development of low TV and high ETCO2 levels during the final induction steps.	Adjust ventilator settings (Pinsp and RR).
obstruction (Respiration, 1)	During documentation steps, an obstruction (e.g., mucus plug) builds in the endotracheal tube. This can be observed by gradually decreasing TV levels, eventually leading to high ETCO2 levels.	Select "Pass suction into endotracheal tube" from troubleshooting menu, complete subroutine.

Table 5.2 (continued): Serious health events, listed in order of their occurrence within the induction task, summary descriptions of the events, and relative properties.

Study Participation

A total of 27 practicing anesthesiologists affiliated with the University of Michigan Department of Anesthesiology participated in this study. The first 11 participants, including 4 females and 7 males aged 26 – 43 years (mean of 30.2 years) completed pilot versions of the study which led to refinements in the experimental design. The remaining 16 participants conducted the study under the refined design; theirs is the only data reported here. These participants included 15 males and 1 female, with an age range of 26-35 years (mean of 29.5 years). One participant was a faculty member, the remaining participants were anesthesiology residents. Five participants were in their first year of residency, 5 were in their second, and 5 were in their third year. All participants had normal or corrected-to-normal vision and hearing, and none had any injuries or conditions that could diminish the sensitivity of the back or arms to vibrotactile stimulation. Each had at least 5 hours of sleep the previous night, and none had consumed more than 2 caffeinated beverages on the day of the experiment.

Procedure

Upon arrival, participants first read through and signed the consent form for the study, then filled out a brief background questionnaire. After a brief explanation of the purpose of the study and the procedure, the experimenter guided participants through an interactive training session on the induction and physiological management tasks in the TAPIS simulator. Next, the three tactile displays were explained and participants were presented with each parameter level and dynamic that they would encounter during the experiment.

After familiarizing participants with the three tactile displays, they then completed a "pre-test" to further train them in interpreting the hybrid display. Because the hybrid display contained components of both the continuous and alarm displays, it was only necessary to complete the pre-test under the hybrid display configuration. The pre-test consisted of 15 separate presentation sequences which consisted of first a full presentation of normal conditions (BP, TV, and ETCO2 parameter at target levels (see Table 5.1) and RR of 15 breaths/minute), then a presentation in which one or multiple changes took place in the parameter levels (BP, TV, ETCO2, and/or RR), and finally a third full presentation with the parameters at these changed levels. Following each presentation sequence, participants had to respond with which parameter(s) changed and in which direction (up or down). The specific changes for each presentation sequence in this pre-test, and the scores for each participant, are listed in Appendix 6. Each participant correctly identified at least 77% of all the changes in the pretest, and all but one identified at least 83% correctly (overall average: 93.1%). Participants had the greatest problems identifying RR changes. This is understandable because, in the isolated presentations of the pre-test, participants lacked the ability to determine changes in respiration rate by the frequency of breaths (i.e., feeling breaths that come faster than every 4 seconds), instead they had to rely solely on how rapidly the inhalation (TV component) and exhalation (ETCO2 component) portions of a single breath were presented.

Following the pre-test, participants completed a final training session which had them complete the induction scenario while concurrently managing BP and respiration parameters under the hybrid display configuration. Dynamics representative of every health event occurred during this session, and participants practiced the instructed responses of pressing the Big Red Button, verbally announcing the symptoms of the event, and then taking the proper corrective actions. Notably, this session gave participants a chance to clearly feel vibration patterns that reflected changes in the simulated patient's RR over time, a dynamic which could not be represented easily in the "pre-test" training exercise. During this final training session, participants were encouraged to watch the visual display of the parameter levels as they changed to get a sense for how these levels and changes were reflected in the vibration patterns of the hybrid display.

The training activities up to this point took approximately 75-90 minutes. After being given a chance to ask any remaining questions, the participants then began the 4 experimental scenarios, each of which was conducted under a different display configuration (see *Experimental Design*, below). Each scenario was completed when all of the anesthesia induction tasks were completed and the participant declared the patient's health to be stable. Immediately following each scenario, participants completed a NASA-TLX worksheet (see Appendix 7) to subjectively rate their experience conducting the required tasks in the scenario they had just completed.

Completion of each scenario and NASA-TLX worksheet took between 10 and 15 minutes. After completing Scenario 4, participants filled out a postexperiment questionnaire (see Appendix 8), which asked them to rank the four display configurations according to a number of attributes and also to provide feedback which could have been used to identify any simulator problems, strategies, or misunderstandings that may have affected the data. For the approximately 2.5 hours required to complete the entire experiment, participants were compensated $200.

Experimental Design

A prospective power analysis found that sufficient statistical power could be achieved with 16 participants (power ≈ .945 when testing at the .05 level of significance to find differences of at least 15 seconds in event detection time between display configurations, and estimating variance for this measure to be 100 s^2). While this power level may appear to be a bit higher than is necessary, a balanced design required that the number of participants be a multiple of 4, and a power analysis under the same assumptions with 12 participants showed power of approximately .765, which was deemed to be too low.

The following balanced incomplete block design (Table 5.3) was employed such that each participant completed the four scenarios (always in order: Scenario 1 – Scenario 4) with a unique order for display configuration. Across the 16 participants, this design allowed each scenario to be completed under each display configuration exactly 4 times. Whenever it was possible to do so, the participants' level of experience as practicing anesthesiologists was balanced in regard to the display configuration employed for Scenario 1. This was done to account for the fact that those with less experience were expected to show a steeper learning curve across scenarios.

	Display configuration for:			
Participant number	**Scenario 1**	**Scenario 2**	**Scenario 3**	**Scenario 4**
1	hybrid	baseline	alarm	continuous
2	continuous	hybrid	baseline	alarm
3	alarm	continuous	hybrid	baseline
4	baseline	alarm	continuous	hybrid
5	hybrid	continuous	alarm	baseline
6	baseline	hybrid	continuous	alarm
7	alarm	baseline	hybrid	continuous
8	continuous	alarm	baseline	hybrid
9	continuous	alarm	hybrid	baseline
10	baseline	continuous	alarm	hybrid
11	hybrid	baseline	continuous	alarm
12	alarm	hybrid	baseline	continuous
13	baseline	hybrid	alarm	continuous
14	continuous	baseline	hybrid	alarm
15	alarm	continuous	baseline	hybrid
16	hybrid	alarm	continuous	baseline

Table 5.3: Balanced design used to assign display configurations to scenarios for each participant.

Measures

The primary within-subjects independent variable for this study was *display configuration* (baseline, alarm, continuous, hybrid). *Experience as a practicing anesthesiologist* (which was classified as first-, second-, third-year resident, or faculty) represented a between-subjects factor.

For some dependent measures, the *type* of event (blood pressure, respiration, or emergency) was handled as an independent variable. Blood pressure and respiratory events (all non-emergency events) were further classified according to two independent variables: *speed of event development* and the amount of visual/attentional *demand during event occurrence*. Table 5.4 below lists how each event was classified according to these two measures. Regarding the speed of event development, "fast" events realized at least 80% of the changes in parameter levels (which, if uncorrected, would result in parameter levels outside of their acceptable ranges) within 15 seconds or 3-4 breaths. "Slow" events involved steady changes in parameter levels, showing 80% of changes occurring over 45 seconds or 10-12 breaths. For example, hypertension events that occurred during laryngoscopy or insertion of the endotracheal tube (the hyper_scope1 or hyper_scope2 events, respectively) were considered "fast": MAP levels rose approximately 30 mmHg over 15-20 seconds. Hypotension that resulted from the administration of the inhaled anesthetic agent isoflurane (hypo_isoflurane) was considered a "slow" event, showing a steady decline of 30 mmHg over roughly 60 seconds.

Events occurring during "high" demand conditions involved induction task steps which required nearly constant visual focus on the patient display and engaged both hands. "Low" demand conditions left the anesthesiologist relatively free to leave the patient display to visually explore physiological data displays. For example, Figure 5.15 depicts an induction task step that imposes high demand. Insertion of the endotracheal tube requires participants to hold the laryngoscope at a precise angle with one hand to allow visualization of the laryngeal opening at the back of the patient's throat, then using the other hand to guide the tube through this opening into the proper place in the patient's airway. "hyper_scope2", the event that was scripted to occur during these insertion steps in Scenario 1, was therefore classified as occurring during a period of high demand. In contrast, the respiratory event "obstruction" was classified as occurring during a low demand condition because the participants were conducting charting/documentation steps, which could be easily interrupted to visually sample physiological displays with virtually no penalty to performance on the induction task.

	Scenario	Event type (parameters involved)	Speed of event development	Demand during event occurrence
hyper_nervous	3	BP	fast	low
shortshallow1	4	Resp (RR, TV, ETCO2)	fast	low
shortshallow2	3	Resp (RR, TV, ETCO2)	slow	low
hypo_propofol	1	BP	slow	low
hypovolemia1	3	BP	slow	high
hyper_scope1	4	BP	fast	high
hyper_vomit	2	BP	fast	low
hyper_scope2	1	BP	fast	high
hypovolemia2	2	BP	slow	high
RMI1	2	Resp (TV)	fast	low
hyperventilation	1	Resp (TV, ETCO2)	fast	high
hypo_isoflurane	4	BP	slow	low
hypercarbia	2	Resp (ETCO2)	slow	high
RMI2	3	Resp (TV)	fast	high
hypoventilation	4	Resp (TV, ETCO2)	slow	high
obstruction	1	Resp (TV, ETCO2)	slow	low

Table 5.4: Classification of blood pressure and respiration events, according to the independent variables *speed of event development* and (visual/attentional) *demand during event occurrence*. Events are listed in chronological order in regard to when they occur within the induction simulation (regardless of scenario).

Dependent measures for this study included *event detection rate*, *event detection time, event correction time*, *physiological management score*, *induction completion time*, *multitask combined performance score*, *subjective workload ratings* according to attributes from the NASA-TLX survey, and *subjective rankings of display configurations* for supporting various attributes related to attention and task management.

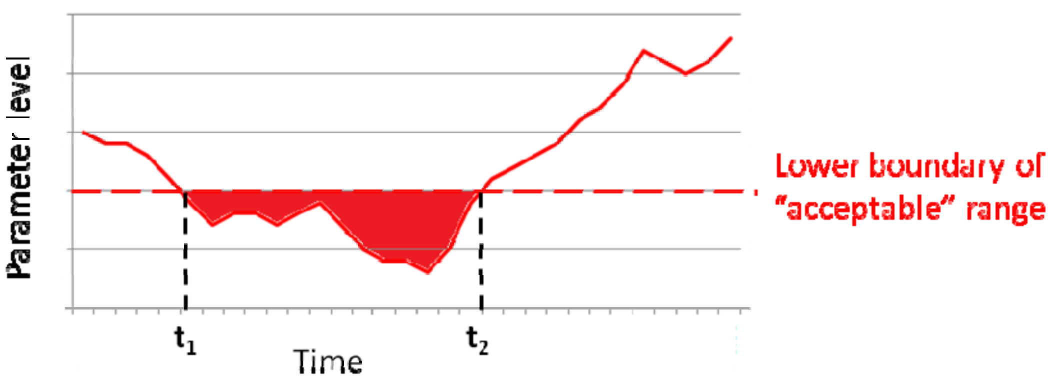
Event detection rate. This measure was defined as the percentage of serious health events that were detected. Events were considered “detected” by a participant when they pressed the Big Red Button (BRB) or another simulation control button (e.g., infusion pump or ventilator setting button), then clearly announced at least one of the symptoms of the event (e.g., “blood pressures are high”, “irregular respiration rate”). During emergency events, if both blood pressure and respiration symptoms were present when the BRB or other control button was pressed, participants were required to announce at least one of each type of symptom. An event was considered to be "missed" if the participant failed to complete these steps before a) the participant declared that they had completed the induction scenario, or b) a later event occurred which masked the abnormal parameter levels and/or dynamics caused by the earlier event. If an event occurred which, due to preexisting scenario conditions, did not result in abnormal levels of a physiological parameter when its full effect was realized, then that event was excluded and rates were calculated according to the remaining events which did have a noticeable effect. For example, if the patient was already very hypertensive due to a previous scenario event when a hypotensive event began, the result of the hypotensive event may offset the hypertension and bring blood pressures back within the acceptable range. In this case, the hypertensive event would be considered "missed" and the hypotensive event would not be considered in the calculation of detection rate.

Event detection time. This was defined as the time elapsed from the first instance that an event causing an abnormal state could be recognized (when it first trended outside of the normal fluctuation range, see Table 5.1) until the BRB or another control button was pressed, provided at least one symptom of the event was correctly identified via verbal report following the button press. If both blood pressure and respiration symptoms were present for an emergency event, the detection time was recorded when the first symptom was announced, regardless of if/when other(s) were reported.

Event correction time. Correction time for an event was defined as the time elapsed from the first instance that an event causing an abnormal state could be recognized until all parameter levels had been brought back within the acceptable range and stabilized. At times, a very high dosage of a drug was administered or ventilator settings were adjusted in a way which caused a parameter to temporarily (re-)enter the acceptable range but "overshoot" this range; in these cases, the parameter level was not yet considered to be stabilized. Note that sometimes participants would not recognize all of the symptoms of a health event and therefore would correct only some of the parameters. When all of the relevant symptoms were verbally reported, the experimenter would reveal the proper corrective actions. Correction times were recorded only after the completion of these actions and the stabilization of all parameter levels. If an event was not fully corrected before being masked by a subsequent event or before the participant declared that they had completed the scenario, the event was considered a “miss” with regard to this measure, and required a data correction procedure (see the *Data Corrections* section).

Physiological management score. This was measured as the sum, over the entire scenario, of the “error areas” for the three critical physiological parameters: areas between the boundaries of the acceptable ranges for each parameter and curves defined by parameter levels over time (see Figure 5.20). Performance was considered worse for higher values of this score, which could be realized with more severe levels of a parameter (greater distance from the acceptable range boundary) and/or longer time spent outside the acceptable range.

Note that “benign” events which caused parameters to temporarily move outside of acceptable ranges were not considered in the physiological management score. For example, following the administration of propofol in the induction task, the patient’s breaths became shallower over a period of 30 seconds until they ceased breathing. Because this is an expected effect of administering propofol, the expected low TV measures during this period did not contribute to “error area”.



$$\text{Physiological management score} \mathrel{+}= \sum_{T=t_1}^{t_2} (\text{acceptable range boundary} - \text{parameter level})$$

Figure 5.20: Calculating physiological management score. A single score was calculated for each scenario as the sum total of all such “error areas” that occurred during the duration of the induction task.

Induction completion time. This measure was defined as the time elapsed from the beginning of the first step in the induction task until the final health event had been corrected, i.e., when all physiological parameters were within acceptable ranges and stabilized. If a participant declared that they were finished with the induction task before all serious health events were fully corrected, these events were considered to be “missed” and a data correction procedure was applied to add an amount of time representative of the missed corrective actions (see the *Data Corrections* section).

Multitask performance score. This score quantified a participant’s performance in each scenario by adding measures of performance for the two primary tasks of induction and physiological management, normalized by the average performance on each task across all participants in the same scenario. Equation 5.1 was used to calculate these scores. Because participants were instructed to give equal priority to completing the induction task as quickly as possible and managing physiological measures as best they could, the two tasks were given equivalent weighting (1/2) in this score. Further, since managing each of the MAP, ETCO2, and TV measures were considered equally important, they were also given equivalent weightings (1/3). Higher scores represent worse multitask performance, and a score of “1” represents performance that is equivalent to the mean performance level across all participants for the given scenario regardless of display configuration.

$$P_{DS} = \left(\begin{array}{l} \frac{1}{2}\left(\frac{1}{3}\left(\frac{MAP\_mgmt(D)}{MAP\_mgmt(Avg_S)} \right) + \frac{1}{3}\left(\frac{ETCO2\_mgmt(D)}{ETCO2\_mgmt(Avg_S)} \right) + \frac{1}{3}\left(\frac{TV\_mgmt(D)}{TV\_mgmt(Avg_S)} \right) \right) \\ + \frac{1}{2}\left(\frac{InductionCompletionTime(D)}{InductionCompletionTime(Avg_S)} \right) \end{array} \right)$$

Equation 5.1: Multitask combined performance score (P_{DS}). *D* = Display configuration (baseline, alarm, continuous, or hybrid); *S* = Scenario (1, 2, 3, or 4) which the participant completed under configuration *D*; *[X]_mgmt(D)* = participant’s physiological management score for parameter [X] in scenario *S* and under configuration *D*; Avg_S = average performance across all participants in scenario *S*.

Subjective workload ratings. Appendix 7 includes a subjective evaluation worksheet for the 6 attributes of workload which are used to calculate the NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988). Following each scenario, participants were asked to rate *mental demand, physical demand, temporal demand, performance, effort,* and *frustration* to describe their experience while conducting the tasks required in the scenario. They placed marks along analog scales that were coded as scores from 0-20 (“10” represented marks at the very center of the scale) for statistical analysis. Participants were told that, while the display configuration may factor into their ratings, they were to focus on the overall experience of conducting the multitask set in the specific scenario. They were encouraged to review surveys from previous scenarios when rating their relative experience in the most recent scenario.

This assessment was meant to be done very quickly between scenarios and some attributes (in particular, mental demand, effort, and frustration) were deemed considerably more important than the others for this study. Therefore the full NASA-TLX procedure was not conducted (which would have involved additional procedures to calculate the relative weightings participants assigned to each attribute in their evaluation process); instead, ratings for each attribute were considered independently.

Subjective rankings of display configurations. Following the completion of all 4 scenarios and NASA-TLX worksheets, participants completed a postexperiment questionnaire (see Appendix 8) which asked them to rank each of the 4 display configurations (from 1 to 4; ties were allowed) along attributes that included *supporting multitask performance*, *minimizing cognitive demand, minimizing annoyance, minimizing unnecessary distraction, maximizing trust, comfort,* and *overall preference*.

Data Corrections

Two emergency health events, inadequate_depth (Scenario 3) and PEA (Scenario 4) were found to greatly impact physiological management scores, to the extent that the scores for the respective scenarios were almost entirely determined by the response to those events. For example, the PEA event involved the patient's blood pressure levels rapidly dropping to nearly zero and spontaneous breathing to rapidly get shallower and then cease, all within about 10 seconds. Participants took an average of about 33 seconds to diagnose this event and perform CPR to correct it. For the roughly 23 seconds in between PEA development and correction, physiological levels for blood pressure and respiration parameters were left far outside their respective acceptable ranges, leading to very large management error area that made all other error for the remainder of the scenarios statistically inconsequential to the physiological management score. As a result, the error due to these two emergency events was not considered in this score (detection and correction times were still analyzed).

When participants missed an event, the effect of this miss on the physiological management score was highly dependent on additional factors, including the speed with which certain induction task steps were completed and pre-existing drug infusion and ventilator settings. For example, the RMI2 event involved the displacement of the endotracheal tube from the trachea into the right mainstem bronchus in the very final steps of the induction task, resulting in tidal volumes about half of normal levels (because only one lung is inflated). This event was considered active until the participant either took the steps necessary to adjust the position of the tube or, following the operating definition of a "miss", declared the patient stable at the completion of the induction task. This introduced a problem in the data analysis because for this event (and some others), the error area associated with the event could be less for participants that missed the event than for those that corrected it because the event could be considered active for a longer period of time in the latter case. To correct for this problem, a data cleaning procedure was used to nullify the actual error area caused by missed events, and to instead add to the physiological management score a value equal to the largest error area associated with the same event from the set of participants that properly corrected the event. For example, 14 of the 16 participants properly repositioned the endotracheal tube to correct the RMI2 event. Of these 14 participants, the largest error area that resulted exclusively from this event was 10621 mL*s for the TV parameter. For the 2 participants who missed the event, the TV error area associated with RMI2 in their data was replaced with the amount 10621, therefore adding this amount to their physiological management score for TV.

A similar rule was applied to the detection and correction times for missed events. Because the experimental design only allowed for one presentation of an event for each combination of independent factor levels, missing an event and therefore not having an associated detection/correction time made it impossible to consider any of the participant's time-based data in a balanced statistical analysis. Therefore, for the time-based analyses, the values entered for the detection and correction times for participants which missed an event were the longest times observed from the set of participants which properly corrected that event. The rationale for this data cleaning procedure was that completely missing an event should be at least as detrimental to the health of the patient as would be the poorest performance in attempting to correct the event. When missed events were active at the end of a scenario, the induction completion time for that participant was recorded as the scenario time when the event was activated plus the longest correction time across all participants.

The last data cleaning procedure involved the events which, due to pre-existing conditions (most notably missing a previous event), did not result in abnormal parameter levels after the full effects of the event were realized. An example of this occurrence was hypotension that followed a missed hypertensive event (e.g., hypo_isoflurane following hyper_scope1 in Scenario 4), or vice-versa. When no corrective actions were taken for either event, the effects of the two events essentially cancelled out, resulting in MAP levels within the acceptable range. Commonly, physiological management scores would be lower in these cases than if both the hypertension and hypotension events had been properly corrected. In these cases, the event that was missed was handled according to the procedures described in the previous two paragraphs. The latter event – the one which was not noticeable because it effectively cancelled out the effects of the missed previous event – was assigned values for detection time, correction time, and physiological management error area that were equal to the average values associated with the same event among participants who conducted the event's scenario under the same display configuration and handled each event in the normal manner (i.e., correcting the hypertensive event, realizing the full effect of the hypotensive event, and then correcting it).

Analysis

Most dependent measures (all the performance-based measures) were analyzed in repeated measures ANOVAs, formulated in the software package PASW Statistics 17. Subjective rankings were analyzed in nonparametric Friedman tests. In analyses where *experience as a practicing anesthesiologist* was found to be an insignificant factor and not part of any significant interaction effect, it was removed from the statistical model. When significant factor effects were found in the repeated measures ANOVAs, post-hoc Fisher's Least Significant Difference tests were used to identify significant differences between factor levels.

RESULTS AND DISCUSSION

The following sections present the results from the statistical analysis, roughly following the order of the hypotheses that were introduced in the Hypotheses section (pages 118-124) and grouped by the major dependent measures: 1) event detection rates, 2) event detection and correction times, 3) physiological management scores, 4) induction completion times, 5) multitask performance scores, 6) subjective workload ratings, and 7) subjective rankings of display configurations. Discussion, as it relates to each hypothesis or interesting finding, follows where appropriate. Following this section, a more general discussion summarizes the most significant findings and their implications.

<u>1) Event Detection Rates</u>

Results

Table 5.5 shows the event detection rates, as percentages, for each participant in a) each scenario and b) under each display configuration. Repeated measures ANOVAs found no difference in detection rate among scenarios, but did find a significant effect of display configuration ($F(3,13) = 4.669$; $p = .020$). Post-hoc Fisher's LSD tests showed that participants performed significantly worse in the baseline configuration (mean detection rate: 79.06%) than in the alarm (97.19%; $p = .004$), continuous (97.16%; $p = .015$), and hybrid configurations (100%; $p = .003$). While the hybrid configuration showed the highest detection rates overall, these rates were not significantly different from those of the alarm or continuous configurations. Experience as a practicing anesthesiologist did not significantly affect detection rates, nor was it a component in any significant interaction effects, and it was therefore not considered in the statistical model.

	a) Scenario			
participant number	1	2	3	4
1	100	75	100	100
2	100	100	60	100
3	75	100	100	25
4	50	100	100	100
5	100	100	100	100
6	100	100	100	100
7	100	100	100	100
8	100	100	40	100
9	75	100	100	80
10	100	100	100	100
11	100	100	100	100
12	100	100	80	100
13	100	100	100	100
14	100	75	100	80
15	100	100	80	100
16	100	100	80	100
Mean	**93.75**	**96.88**	**90**	**92.81**

	b) Display configuration			
participant number	baseline	alarm	continuous	hybrid
1	75	100	100	100
2	60	100	100	100
3	25	75	100	100
4	50	100	100	100
5	100	100	100	100
6	100	100	100	100
7	100	100	100	100
8	40	100	100	100
9	80	100	75	100
10	100	100	100	100
11	100	100	100	100
12	80	100	100	100
13	100	100	100	100
14	75	80	100	100
15	80	100	100	100
16	100	100	80	100
Mean	**79.06**	**97.19**	**97.19**	**100**

Table 5.5: Detection rates (%) for each participant by a) Scenario and b) Display configuration. Display configuration significantly affected detection rate, with the baseline mean (79.06%) significantly lower than those for the three tactile display configurations.

Table 5.6 shows detection rates for each display configuration and each individual scenario event (listed in chronological order by scenario) and a summary of event detections broken down by *type* (blood pressure, respiration, or emergency), the *speed of event development*, and *demand during event occurrence*. An analysis comparing detection rates among blood pressure, respiration, and emergency events found no significant differences due to the type of event. However, because emergency events could be detected by changes or abnormal levels in either blood pressure or respiration measures, separate analyses were conducted to compare blood pressure (only) events to respiration (only) events. These analyses did show a significant difference: detection rates across all display configurations for blood pressure events (96.88%) were higher than those for respiration events (90.63%; $F(1,15) = 6.000$; $p = .027$). The *speed of event development factors, demand during event occurrence*, and any interaction effects were not found to significantly affect detection rate.

	Event (in chronological order)	baseline	alarm	continuous	hybrid	Mean
Scenario 1	hypo_propofol	100	75	100	100	93.75
	hyper_scope2	100	100	100	100	100
	hypervent	75	100	75	100	87.5
	pneumothorax	75	100	100	100	93.75
	obstruction	100	100	100	100	100
Scenario 2	hyper_vomit	50	100	100	100	87.5
	hypovolemia2	100	100	100	100	100
	RMI1	100	100	100	100	100
	hypercarbia	100	100	100	100	100
	anaphylaxis	100	100	100	100	100
Scenario 3	hyper_nervous	100	100	100	100	100
	shortshallow2	0	100	100	100	75
	hypovolemia1	100	100	100	100	100
	inadequate_depth	50	100	100	100	87.5
	RMI2	75	100	75	100	87.5
Scenario 4	shortshallow1	50	75	100	100	81.25
	PEA	100	100	100	100	100
	hyper_scope1	75	100	100	100	93.75
	hypo_isoflurane	100	100	100	100	100
	hypoventilation	75	100	100	100	93.75

(Display configuration: baseline, alarm, continuous, hybrid)

	baseline	alarm	continuous	hybrid	Mean
ALL Blood Pressure events	**90.63**	**96.88**	**100**	**100**	**96.88**
BP events: fast-developing	81.25	100	100	100	95.31
BP events: slow-developing	100	93.75	100	100	98.44
BP events: high demand	93.75	100	100	100	98.44
BP events: low demand	87.5	93.75	100	100	95.31
ALL Respiration events	**71.88**	**96.88**	**93.75**	**100**	**90.63**
Resp events: fast-developing	75	93.75	87.5	100	89.06
Resp events: slow-developing	68.75	100	100	100	92.19
Resp events: high demand	81.25	100	87.5	100	92.19
Resp events: low demand	62.5	93.75	100	100	89.06
ALL Emergency events	**81.25**	**100**	**100**	**100**	**95.31**

Table 5.6: Detection rates in each display configuration for each individual event, and events broken down by independent measures of type (blood pressure, respiration, emergency), speed of development (fast- or slow-developing), and demand during event occurrence (high or low).

Discussion

H1: Participants will show higher detection rates for critical health events when equipped with any of the three vibrotactile displays, when compared to a baseline (i.e., visual only) display configuration.
H2: Participants will show higher event detection rates when equipped with the alarm and hybrid displays than with the continuous display.

H1 is confirmed by these results, but H2 is not. Event detection rate improved nearly 20% with the tactile displays (97 – 100%) compared to without them (79%). The fact that not even a single event was missed in the hybrid display configuration (the only configuration where this was the case) suggests that this display may best support event detection. However, likely due in part to the low number of events overall, and potentially to a ceiling effect, this advantage did not reach significance. More events and/or greater scenario difficulty may have led to significant differences in event detection rates between tactile display configurations.

Contrary to expectations, performance was equivalent with the continuous display and with the alarm display (and statistically, with the hybrid display as well). This suggests that the concerns regarding the higher tendency for tactile change blindness and perceptual masking effects with the continuous display may be unfounded.

Respiration events tended to be missed statistically more often than blood pressure events. Though the interaction effect did not reach significance, an interesting pattern worth noting is that only the baseline and continuous configurations showed a difference between the two events (every event was detected in the hybrid configuration, and the alarm configuration showed equivalent detection rates for both types of events). This might suggest that the “beat” pulse – a component of both the alarm and hybrid displays – was especially useful in supporting the detection of respiration events. Baseline detection rates suggest that the respiration events may have been harder overall to detect, but these results may be dominated by this configuration’s very low detection rates for the two respiration events that occurred while the patient was still spontaneously breathing: shortshallow1 (baseline detection rate: 50%) and shortshallow2 (0%).

In the baseline case, these low rates might be explained by the fact that respiration activity is not commonly monitored very closely – at least not according to visually-displayed respiration data – until the patient is under mechanical ventilation. The fact that the inadequate_depth emergency event (which involved respiration activity when it should not have been expected, prior to the ventilator being turned on) also showed very low detection rates in this configuration (50%) may support this idea. In fact, of the two participants who detected and correctly diagnosed the inadequate_depth event under the baseline configuration, one initially noticed the subtle "bucking" of the patient and the other did not detect the event until after the ventilator was turned on. Neither paid much attention to the respiration parameters prior to the ventilator being turned on, suggesting that a tactile display related to respiration parameters may be especially beneficial for detecting events that occur prior to mechanical ventilation.

Though it was not stated in the hypotheses, there was an expectation that performance would be generally worse in earlier scenarios than later scenarios, because participants would improve with more experience in the simulator. In terms of event detection, this was not the case: no difference was found between the scenarios; if anything, performance was slightly worse in the later two scenarios than the first two. However, this might also reflect the low detection rates for the three events that most plagued the baseline configuration: shortshallow1, shortshallow2, and inadequate_depth. The fact that a difference between scenarios was not found for the event detection rate measure serves as evidence that the scenarios were roughly equivalent in difficulty (or that the later ones involved increased difficulty that effectively balanced performance benefits due to experience), and thus comparison of performance among display configurations and across all scenarios is appropriate.

2) Detection and Correction Times

Results

Average detection and correction times across all events in each scenario/display configuration were calculated for each participant. Repeated measures ANOVAs of these averages found *display configuration* to significantly affect detection time (DT: $F(3,13) = 46.418$; $p < .001$) and correction time (CT: $F(3,13) = 23.253$; $p < .001$). *Experience as a practicing anesthesiologist* was not a significant factor in either analysis. Post-hoc Fisher's LSD tests for detection time found that the baseline display configuration (mean DT: 56.4 s) featured significantly longer detection times than did the alarm configuration (DT: 28.1 s; $p < .001$), the continuous configuration (DT: 26.8 s; $p = .001$), and the hybrid configuration (DT: 14.0 s; $p < .001$) (see Figure 5.21). The hybrid configuration also showed significantly faster detection times than the alarm ($p < .001$) and continuous ($p = .003$) configurations. Correction times showed a similar pattern, with significantly longer correction times in the baseline configuration (mean CT: 104.8 s) than the alarm (CT: 72.1 s; $p = .001$), continuous (CT: 64.7 s; $p = .001$), and hybrid (CT: 46.0; $p < .001$) configurations. The correction times for the hybrid configuration were also significantly faster than those for the alarm ($p < .001$) and continuous ($p = .001$) configurations. The alarm and continuous display configurations did not significantly differ for either measure.

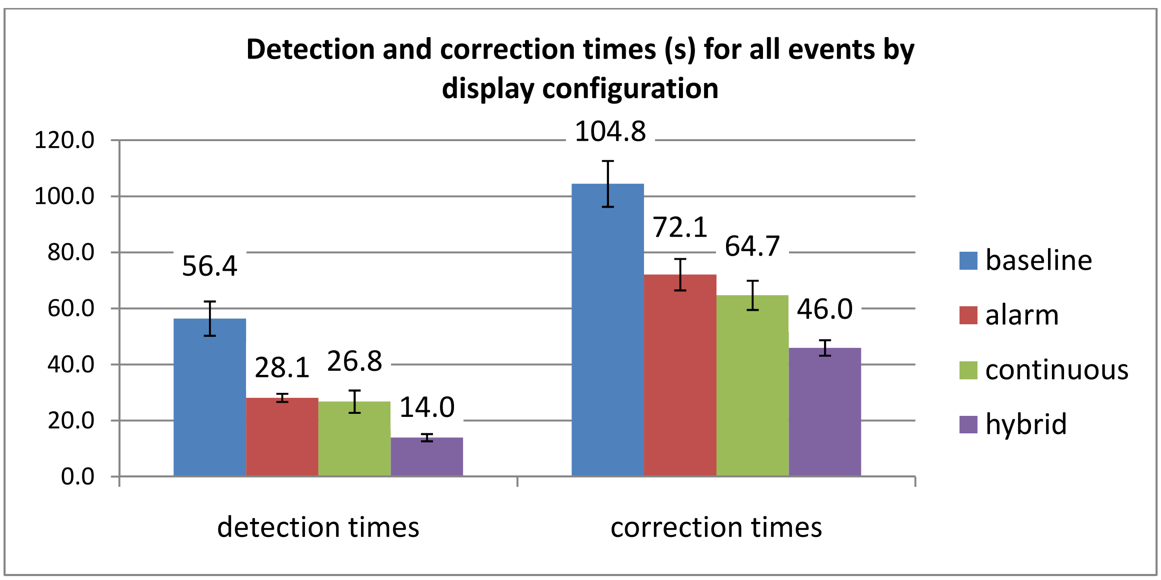


Figure 5.21: Detection and Correction times for all events, by display configuration. Error bars represent standard error. Baseline configurations showed significantly longer DT and CT, and hybrid configurations showed significantly shorter DT and CT, than all other display configurations.

A separate set of analyses was run to determine the effect of *scenario* on detection and correction times, regardless of display configuration (see Figure 5.22). It was expected that performance would improve somewhat in later scenarios, when participants had more experience with the simulator and required tasks. However, detection time was not found to be significantly affected by *scenario*. For correction time, there was at best a trend suggesting an effect ($F(3,13) = 2.717$; $p = .088$), with slightly faster CT in Scenario 4.

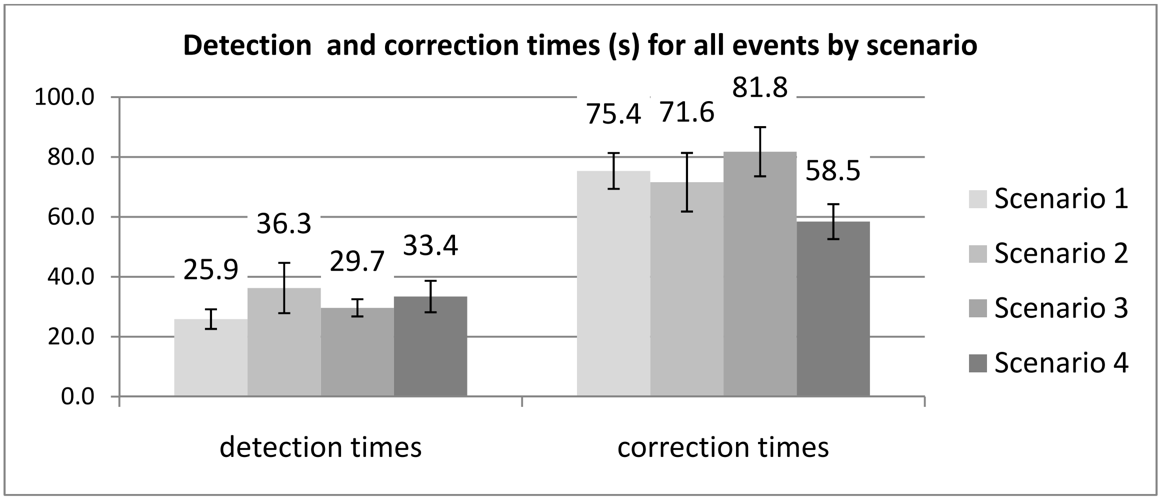
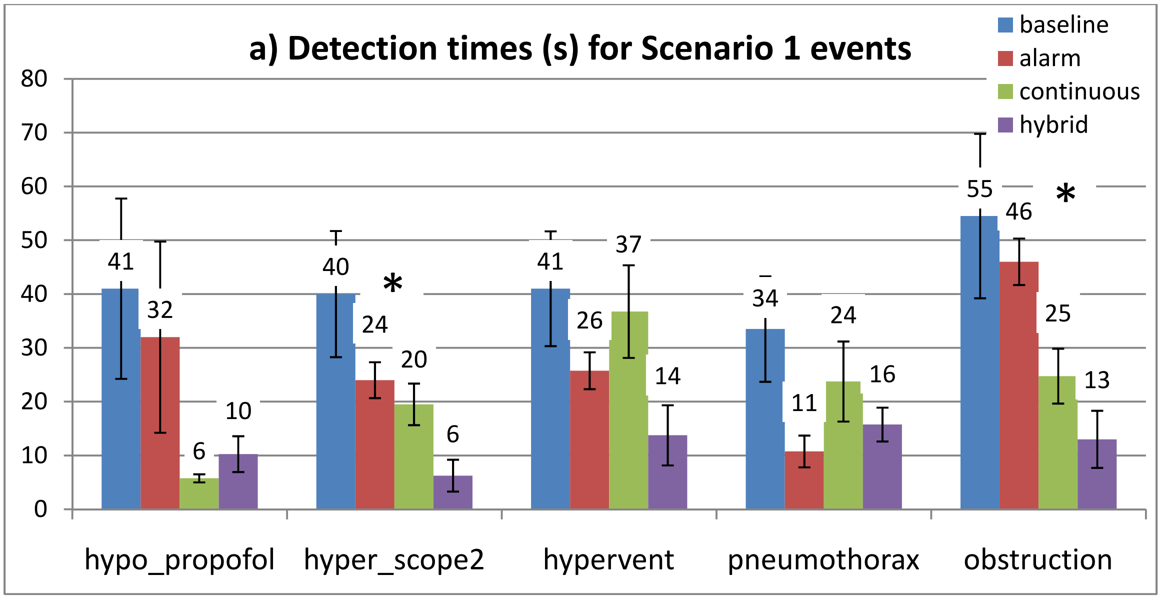
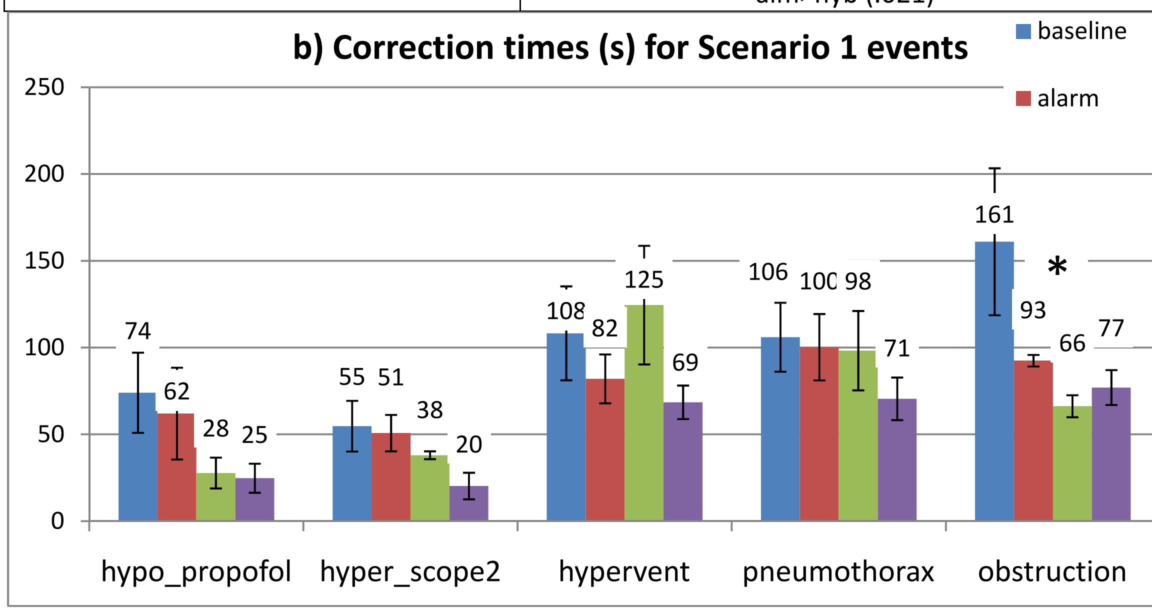


Figure 5.22: Detection and Correction times for all events, by scenario. Error bars represent standard error. CT in Scenario 4 was significantly faster than in Scenario 1 and marginally faster than in Scenario 3.

Figures 6.23-6.26 show (a) detection times and (b) correction times for events in each scenario. Tables immediately below each figure highlight events for which the measure was significantly affected by *display configuration*, and any significant differences between means, as determined by Fisher's LSD post-hoc tests.

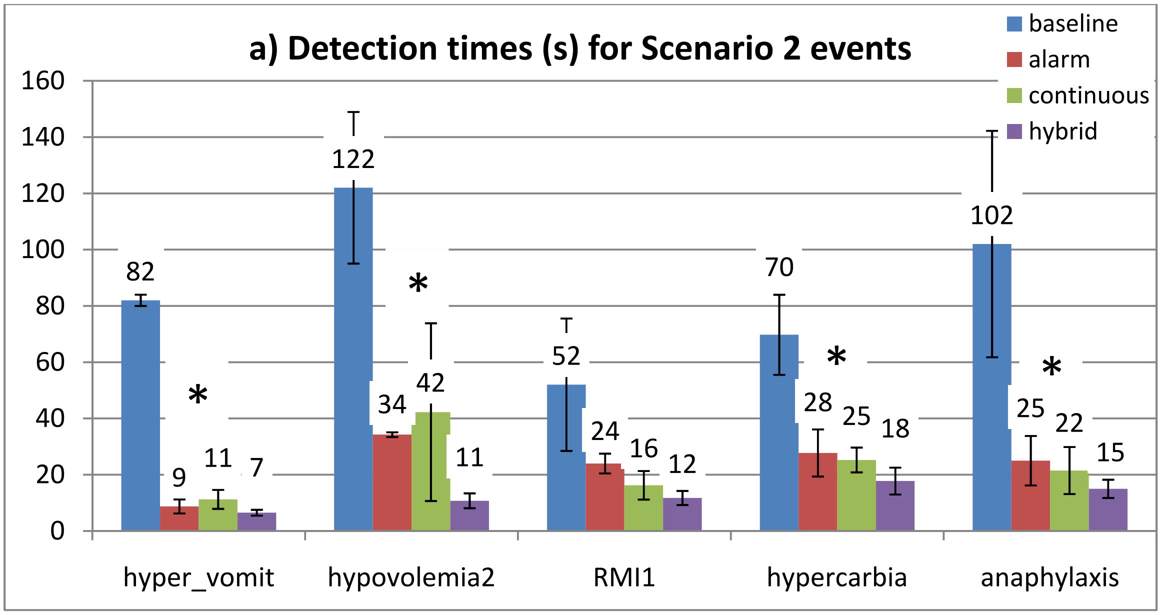


Significant effect of display config on DT:	Significant post-hoc comparisons (p-values)
hyper_scope2 (F(3,12) = 4.500; p = .025)	base>cont (.047); base>hyb (.003)
obstruction (F(3,12) = 4.743; p = .021)	base>cont (.033); base>hyb (.006); alm>hyb (.021)

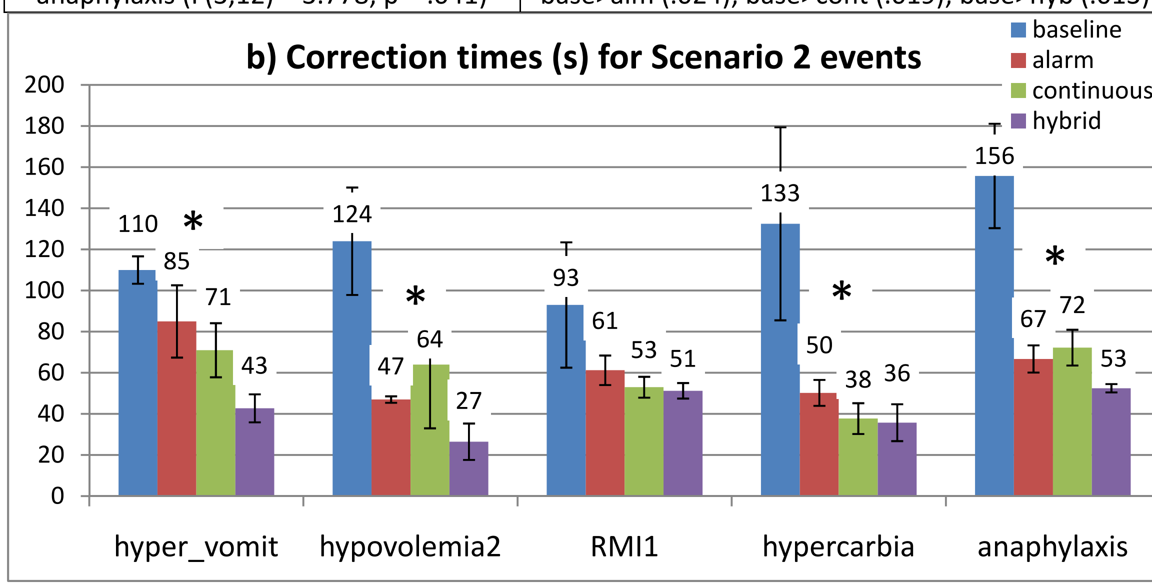


Significant effect of display config on CT:	Significant post-hoc comparisons (p-values)
obstruction (F(3,12) = 3.730; p = .042)	base>alm (.048); base>cont (.010); base>hyb (.006)

Figure 5.23: a) Detection times and b) Correction times for events in Scenario 1. Error bars represent standard error.

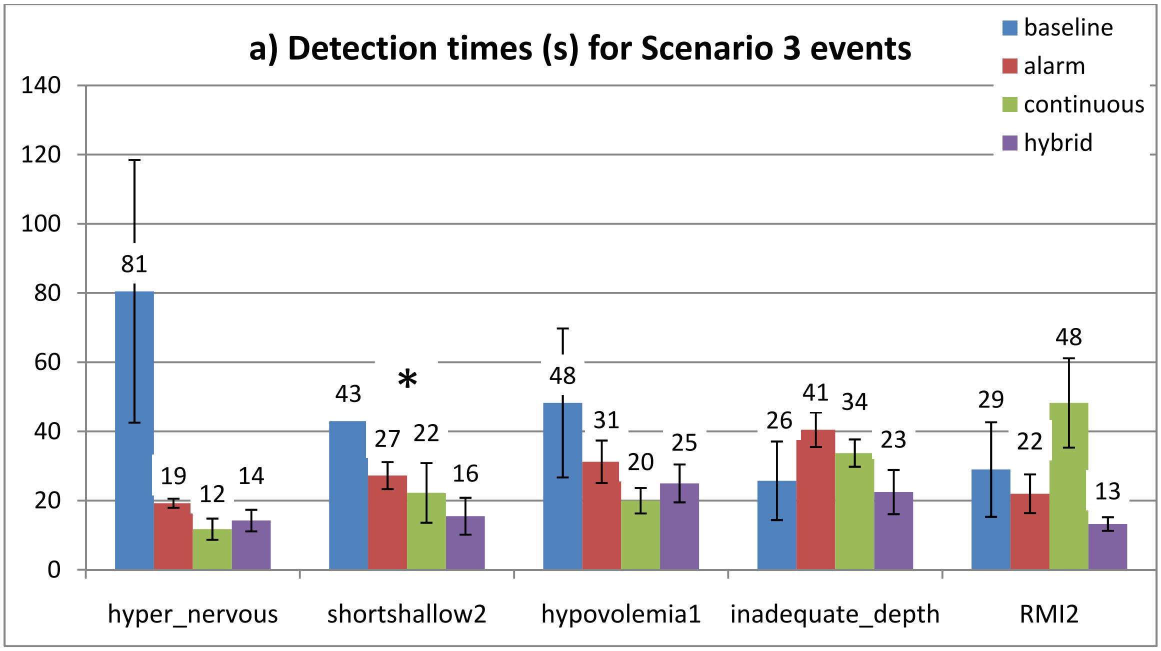


Significant effect of display config on DT:	Significant post-hoc comparisons (p-values)
hyper_vomit (F(3,12) = 236.41; p < .001)	base>alm (.000); base>cont (.000); base>hyb (.000)
hypovolemia2 (F(3,12) = 5.398; p = .014)	base>alm (.011); base>cont (.019); base>hyb (.003)
hypercarbia (F(3,12) = 7.001; p = .006)	base>alm (.006); base>cont (.004); base>hyb (.001)
anaphylaxis (F(3,12) = 3.778; p = .041)	base>alm (.024); base>cont (.019); base>hyb (.013)

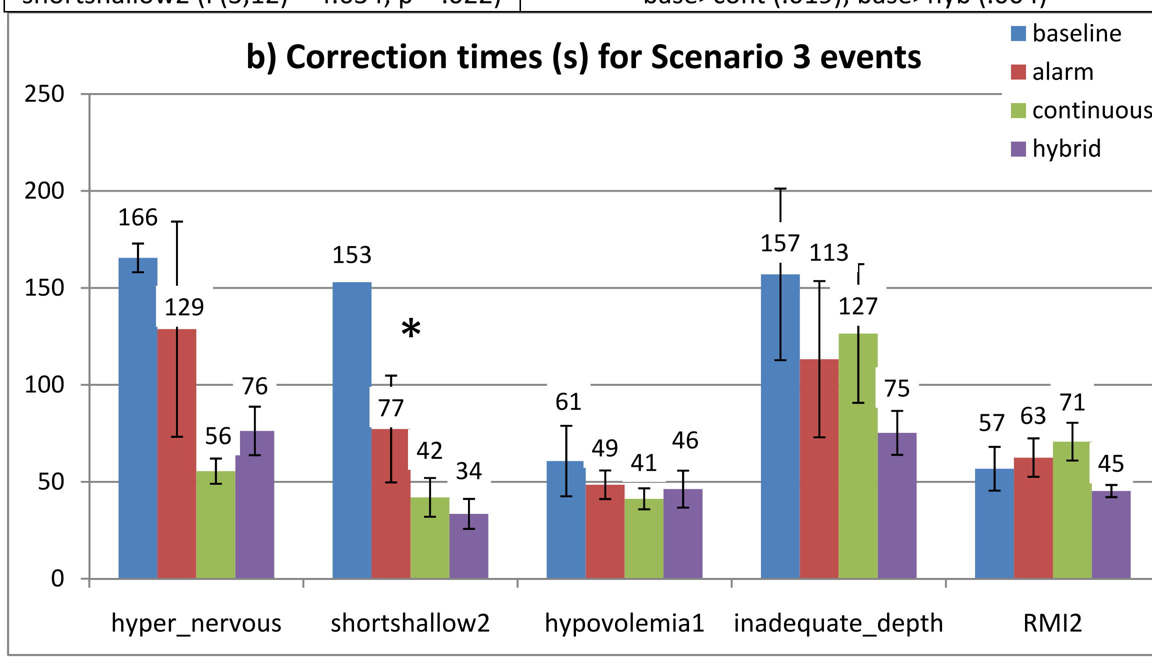


Significant effect of display config on CT:	Significant post-hoc comparisons (p-values)
hyper_vomit (F(3,12) = 5.495; p = .013)	base>cont (.040); base>hyb (.002); alm>hyb (.028)
hypovolemia2 (F(3,12) = 4.089; p = .033)	base>alm (.022); base>hyb (.006)
hypercarbia (F(3,12) = 3.567; p = .047)	base>alm (.035); base>cont (.018); base>hyb (.016)
anaphylaxis (F(3,12) = 11.363; p = .001)	base>alm (.001); base>cont (.001); base>hyb (.000)

Figure 5.24: a) Detection times and b) Correction times for events in Scenario 2. Error bars represent standard error.

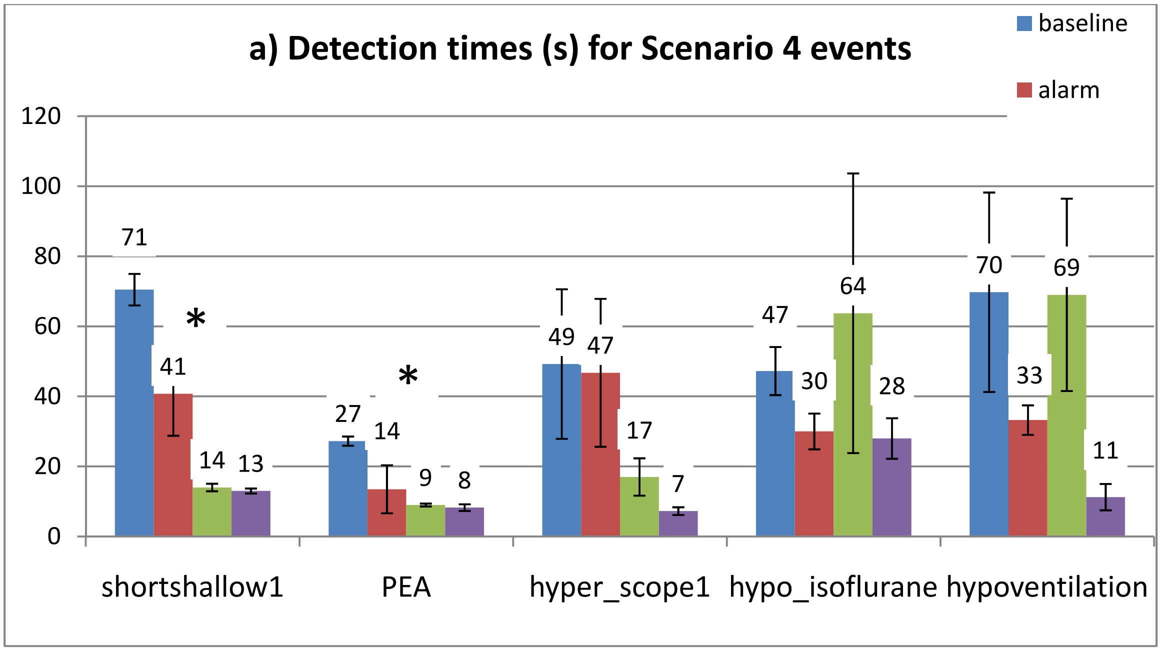


Significant effect of display config on DT:	Significant post-hoc comparisons (p-values)
shortshallow2 (F(3,12) = 4.634; p = .022)	base>cont (.019); base>hyb (.004)

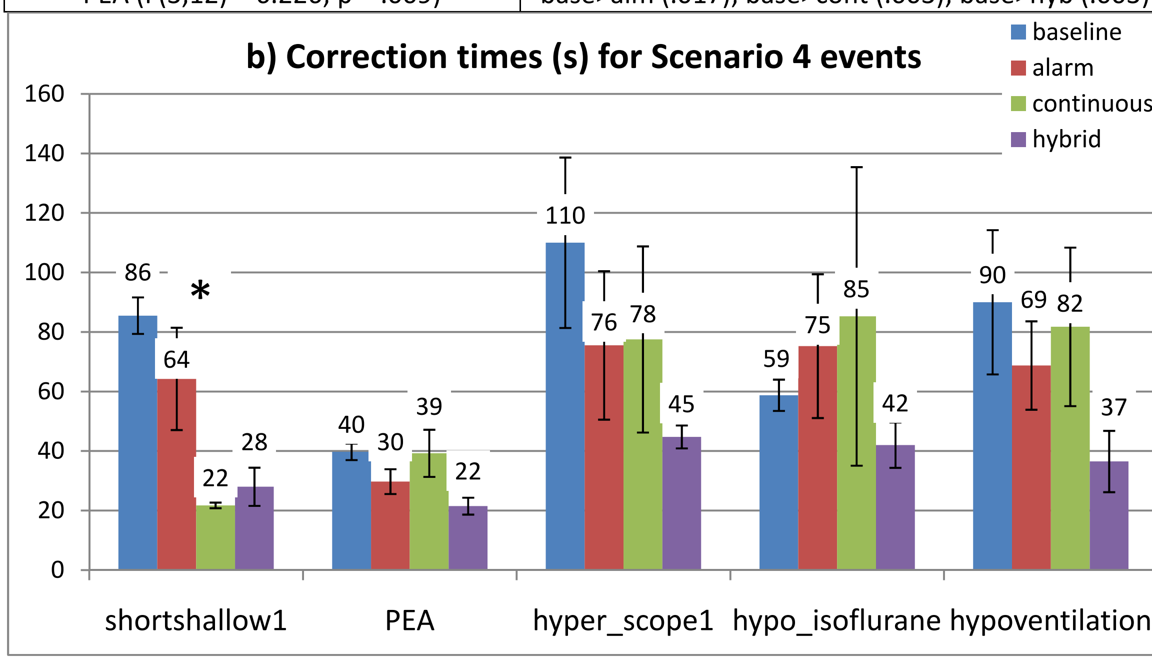


Significant effect of display config on CT:	Significant post-hoc comparisons (p-values)
shortshallow2 (F(3,12) = 12.93; p < .001)	base>alm (.004); base>cont (.000); base>hyb (.000)

Figure 5.25: a) Detection times and b) Correction times for events in Scenario 3. Error bars represent standard error.



Significant effect of display config on DT:	Significant post-hoc comparisons (p-values)
shortshallow1 (F(3,12) = 17.92; p < .001)	base>alm (.007); base>cont (.000); base>hyb (.000); alm>cont (.012); alm>hyb (.010)
PEA (F(3,12) = 6.226; p = .009)	base>alm (.017); base>cont (.003); base>hyb (.003)



Significant effect of display config on CT:	Significant post-hoc comparisons (p-values)
Shortshallow1 (F(3,12) = 9.760; p = .002)	base>cont (.001); base>hyb (.001); alm>cont (.009); alm>hyb (.021)

Figure 5.26: a) Detection times and b) Correction times for events in Scenario 4. Error bars represent standard error.

Figure 5.27a shows detection times, and Figure 5.27b shows correction times, for all blood pressure and respiration events according to the *speed of event development*. Note that emergency events were not included in these analyses, since they didn't strictly adhere to the development timeframes of the other events. Regarding detection time, a repeated measures ANOVA found a significant effect of *display configuration* (mirroring the effect found in the analysis of all events, see Figure 5.21), and a trend suggested an effect of *speed of event development* ($F(1,15) = 3.102$; $p = .099$), with slightly faster times for fast-developing events (overall mean DT: 28.6 s) than for slow-developing events (36.3 s). *Type* of event (blood pressure vs. respiration) did not significantly affect DT, nor did any interaction effects.

The analysis of correction times (Figure 5.27b) also found an effect of *display configuration*, but *speed of event development* did not affect measures of CT. However, a significant interaction effect was found between *type* and *speed of event development* ($F(1,15) = 10.626$; $p = .005$). Post-hoc analysis of this interaction found that while no difference was found between fast-developing blood pressure and respiration events, slow-developing blood pressure events were corrected faster (mean CT: 56.8 s) than slow-developing respiration events (mean CT: 77.2 s; $p = .006$). The interaction effect also showed that within blood pressure events, fast-developing events (mean CT: 75.4 s) took longer to correct than did their slow-developing counterparts (56.8 s; $p = .015$).

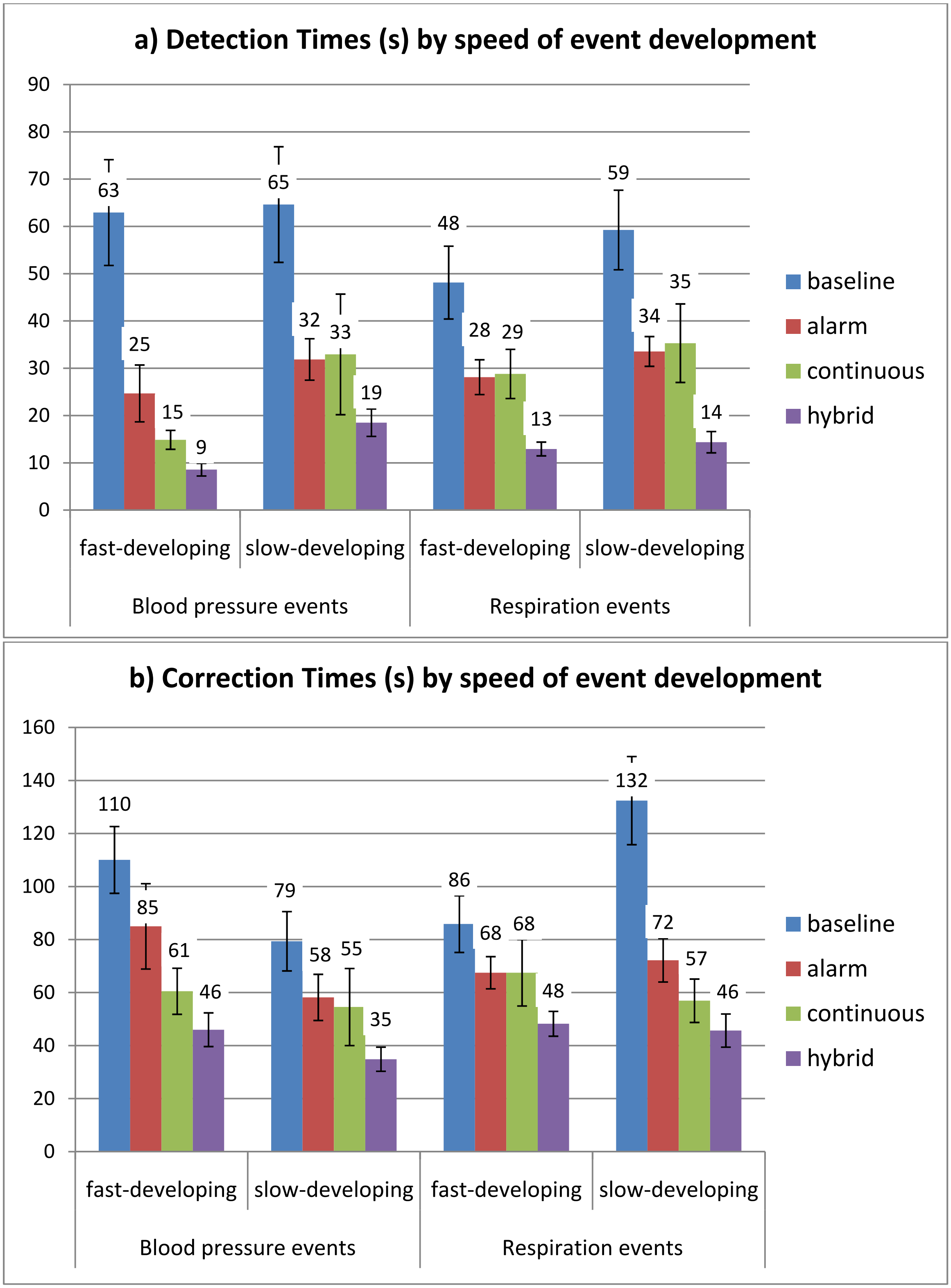


Figure 5.27: a) Detection times and b) Correction times for blood pressure and respiration events, by speed of event development. Error bars represent standard error.

Finally, Figures 6.28a and 6.28b show detection and correction times, respectively, for all blood pressure and respiration events broken down by *demand during event occurrence*. Again, emergency events were not included in these analyses. For detection times, *display configuration* was again found to be a significant factor, but no other main or interaction effects reached significance.

The analysis of correction times, however, showed a significant effect of *display configuration* and a significant interaction effect of *display configuration* by *demand* ($F(3,13) = 4.005$; $p = .032$). Looking into this interaction, it appears *display* significantly affects CT in both high ($F(3,13) = 11.907$; $p < .001$) and low ($F(3,13) = 11.893$; $p < .001$) demand conditions, but in different ways. The most notable differences in post-hoc findings between the two demand levels concern the relative effects of the tactile display configurations: during high demand conditions, alarm CT (60.7 s) and continuous CT (mean CT: 66.9 s) were faster than baseline CT (92.1 s; $p = .005$ and $p = .037$, respectively), but did not differ from each other. Hybrid CT (40.5 s) represented an improvement over all other display configurations under high demand conditions (compared to baseline: $p < .001$; alarm: $p = .002$; and continuous: $p = .002$). Under low demand conditions, baseline CT (mean CT: 112.6 s) was again significantly longer than alarm CT (80.8 s; $p = .036$), continuous CT (52.8 s; $p < .001$), and hybrid CT (46.9 s; $p < .001$). However, in contrast to during high demand conditions, both the continuous and hybrid displays showed faster CT than did the alarm display ($p = .043$ and $p = .003$, respectively), and the continuous and hybrid CTs did not differ.

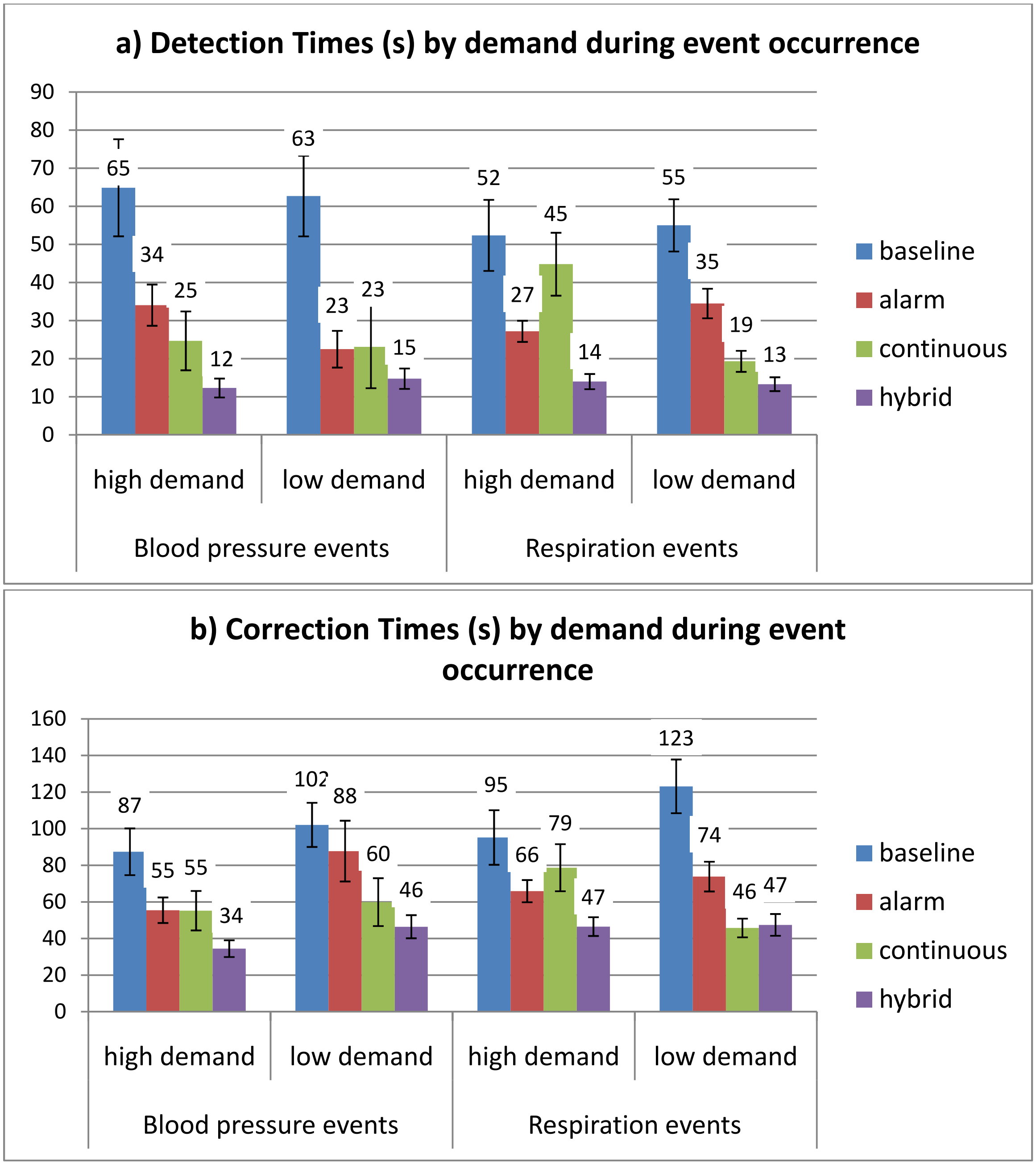


Figure 5.28: a) Detection times and b) Correction times for blood pressure and respiration events, by demand during event occurrence. Error bars represent standard error.

Discussion

Detection and correction times were analyzed to see if any differences existed across scenarios, and to test whether a general improvement could be found across scenarios. The fact that only a trend was observed toward improvements in correction time (but not detection time) is further evidence (along with the lack of effect of *scenario* on event detection rate) that performance did not significantly improve throughout the scenarios for the physiological monitoring task, however, this could again be due to a ceiling effect.

H3: Participants will detect events earlier in their development when equipped with any of the three vibrotactile displays, when compared to the baseline display configuration.
H4: Participants will detect these events earlier when equipped with the continuous and hybrid displays than with the alarm display.

H3 is confirmed, and H4 is partially confirmed in the case of the hybrid display. Detection times were between two and four times as fast with the tactile display configurations (14 – 28 s) than with the baseline configuration (56 s). Detection times for events with the alarm configuration (28 s) and those with the continuous configuration (27 s) were very similar, but the hybrid display showed a large, and significant, improvement in detection time (14 s). This relationship between the four display configurations remained fairly consistent for each individual event: the baseline condition never showed significantly shorter detection times than any other display configuration and frequently showed significantly longer detection times than the other configurations. The hybrid configuration never showed significantly longer detection times than any other configuration, and these times were significantly lower than at least some of the other configurations most often.

The largest DT differences between the baseline and other configurations were found for events which might be considered more unexpected given their place in the simulation: obstruction, hypercarbia, and anaphylaxis, which occurred near the very end of induction when the patient was otherwise relatively stable; and hypovolemia2, which took effect after the endotracheal tube was installed, at a period when blood pressure was otherwise relatively stable. One other event that showed much longer DTs for the baseline configuration was hyper_vomit; this could be due to the fact that the unusual event of vomit in the airway (this was the first of two scenarios that involved this event), and the steps required to address it distracted the participants, such that in the baseline configuration they felt the need to quickly complete the steps required to suction the airway and subsequently forgot to check on the patient's physiological response to the event.

One possible explanation for the surprising finding that the continuous display did not show an improvement over the alarm display is that participants may have tended to ignore the continuous display when it communicated normal and intermediate levels of each parameter. This could be because these levels were less distinguishable in the continuous display than they were in the hybrid display. Because changes that took place at these intermediate levels tended to be the earliest clues of "trending" parameter levels, a clearer ability to distinguish these levels may account for the improved performance in the hybrid configuration that was not observed in the continuous configuration.

The relative benefits of the alarm and continuous displays can be seen to change frequently between individual events, though rarely showing significant differences. A consistent pattern for this relationship could not be identified between factor levels for the major independent measures (*type, speed of event development, or demand during event occurrence*), nor for any interactions between those measures. However, the individual events where the continuous display outperformed the alarm display – obstruction and shortshallow1 – were events that should have given an advantage to the continuous display. Obstruction was a slowly-developing event characterized by a decrease in TV levels when the TV levels should be expected to be stable and little else was required from the induction task, thus the "trending" phase in TV could be more noticeable with the continuous display, and once noticed was clearly diagnostic of a potential problem, hence the earlier response. Shortshallow1 was immediately noticeable with the continuous display because the RR was fast and TV levels were slightly low for a period of time before becoming low enough to activate the low TV alarm. Similarly, the ETCO2 levels did not get low enough to activate an alarm for several breaths.

H5: The performance benefits observed with the continuous and hybrid displays, in terms of earlier event detection, will be greater (i.e., earlier detection) for slowly-developing than for quickly-developing events.

H5 is not confirmed, since a significant interaction effect between *display configuration* and *speed of event development* was not found. This lack of interaction effect, however, may also reflect the consistent benefit that the hybrid display showed over all other configurations for nearly every event, regardless of the conditions that described the event.

H6: Participants will complete corrective actions in response to these events faster when equipped with any of the three vibrotactile displays, when compared to the baseline display configuration.

H7: Participants will complete corrective actions faster when equipped with the continuous and hybrid displays than with the alarm display.

The significant main effects for the correction time measure tended to be very similar to those for the detection time measure. The hybrid display showed significantly better, and the baseline configuration showed significantly worse CTs when compared to every other configuration. Looking into the individual events, a few key differences from the main findings for DT are worth noting. First, some of the events showed considerably larger differences between the baseline configuration and the other configurations for CT than for DT. These tended to be events, such as obstruction, where an abnormal level in one parameter (TV) was noticed and corrected, but abnormal levels in another factor that developed slightly later (ETCO2) were not noticed for a longer period of time, and thus the CT values were especially prolonged in the baseline case, when the event was assumed to be resolved after addressing only the first symptom. Other events that showed this pattern, such as hyper_scope1 and hypercarbia, tended to occur during stretches of the induction task that could be described as extended high-demand periods. Because the full effects of corrective actions for each of these events took time to develop, participants tended to take the actions and then return to the induction task. Frequently, the initial corrective actions were insufficient to fully resolve these events (the remifentanil infusion rate was not set high enough, and the respiration rate ventilator setting was not increased enough). In tactile display configurations, participants received faster feedback related to the failure of the corrective action to fully correct the event, while in the baseline configuration, participants tended to neglect checking on the progress of their corrective action for considerably longer periods of time.

Another difference between the findings for the DT and CT measures may also have been related to the issue of feedback following a corrective action. Some of the blood pressure events, such as hyper_scope2, hyper_vomit, and hyper_nervous, tended to show a large difference in DT between the baseline and alarm configurations, while the CT measures showed this difference to shrink considerably. In effect, the alarm configuration *hurt* performance, relatively speaking, in the resolution of events after they had been detected. This might be explained by the tendency for participants to ignore parameter levels until alarm signals were present, and the frequency of the presentation of blood pressure alarm signals under the alarm configuration. Modeled after many physiological alarms, the tactile signal would vibrate once when the acceptable range threshold was passed by a parameter level, but then would not vibrate again for 20 seconds as long as the parameter level stayed above this threshold. So in the cases of these blood pressure events, participants may detect the event, then start an infusion or adjust the active infusion rate, and then return to the induction task while the changes took effect. If this infusion rate was not high enough to bring MAP levels back within the acceptable range, participants in the alarm configuration would not be notified of this fact for roughly 20 seconds. This delay in feedback may account for the relatively poorer performance in regard to the CT measure.

This "underdosing" behavior, and the lack of effective feedback for MAP levels in the baseline and alarm configurations, might be behind the significant interaction effect found between *type* and *speed of development* for the CT measure. This interaction showed that for blood pressure events, CTs were longer for fast-developing events than for slow-developing events. Closer inspection of this seemingly counterintuitive finding shows that it appears only the baseline and alarm configurations led to appreciably different CTs between fast-developing and slow-developing events. Because all of the fast-developing blood pressure events were cases of hypertension, the proper corrective action required starting a remifentanil infusion. Interestingly, the potency of remifentanil, in terms of how much it lowers BP (-15 mmHg per 10 mL/hr), was half that of its counterpart phenylephrine (+30 mmHg per 10 mL/hr). Even though participants were made well aware of this fact in training procedures, the tendency to program round numbers (such as 10) as infusion rates made the likelihood of underdosing much greater for the fast-developing, i.e., hypertension events.

H12: The demand imposed by scenario conditions will affect detection and correction times in a way that does not vary due to display configuration.

H12 is rejected based on the finding of a significant interaction effect for the CT measure between *display configuration* and *demand during event occurrence*. Breaking down this interaction effect, the hybrid display very consistently resulted in fast CT measures, significantly faster than in both the baseline and alarm configurations for both demand levels. The continuous display, on the other hand, shows CTs that are significantly better than in the alarm configuration only in low demand conditions; in high demand conditions the CTs are considerably worse.

These findings suggest that performance with the hybrid display may be more resilient to changes in workload than the continuous display. This is not entirely unexpected, because one of the designed benefits of the hybrid display was the ability to "tune out" the signal more completely during high demand conditions and allow the beat pulses – those which were not a component of the continuous signal – to effectively capture attention when significant changes in parameter levels took place.

3) Physiological Management Scores

Results

Because the ranges and scales for each of the three primary physiological measures of interest (MAP, ETCO2, and TV) were very different, physiological management scores for each measure were analyzed separately. *Experience as a practicing anesthesiologist* did not show a significant main effect nor did it contribute to any significant interaction effects for any of the analyses of physiological management score. Therefore, this factor was removed from each statistical model.

Figure 5.29 shows the physiological management scores according to each parameter: a) MAP, b) ETCO2, and c) TV. Repeated measures ANOVAs found significant effects of *display configuration* for MAP ($F(3,13) = 7.791$; $p = .003$) and for TV ($F(3,13) = 4.861$; $p = .018$), but not for ETCO2. Post-hoc comparisons showed that MAP management was significantly worse in the baseline configuration (mean error area: 2688 mmHg*s) than in the alarm (1337 mmHg*s; $p = .005$), continuous (1121 mmHg*s; $p = .005$) and hybrid (704 mmHg*s; $p < .001$) configurations. Additionally, the hybrid configuration showed significantly less error area than the alarm configuration ($p = .050$) and a trend toward an improvement over the continuous configuration ($p = .055$). TV management was significantly worse in the baseline configuration (mean error area: 15075 mL*s) than in the alarm (6941 mL*s; $p = .036$) and hybrid (4820 mL*s; $p = .012$) configurations, and a trend toward improvement over the continuous configuration (7192 mL*s; $p = .095$).

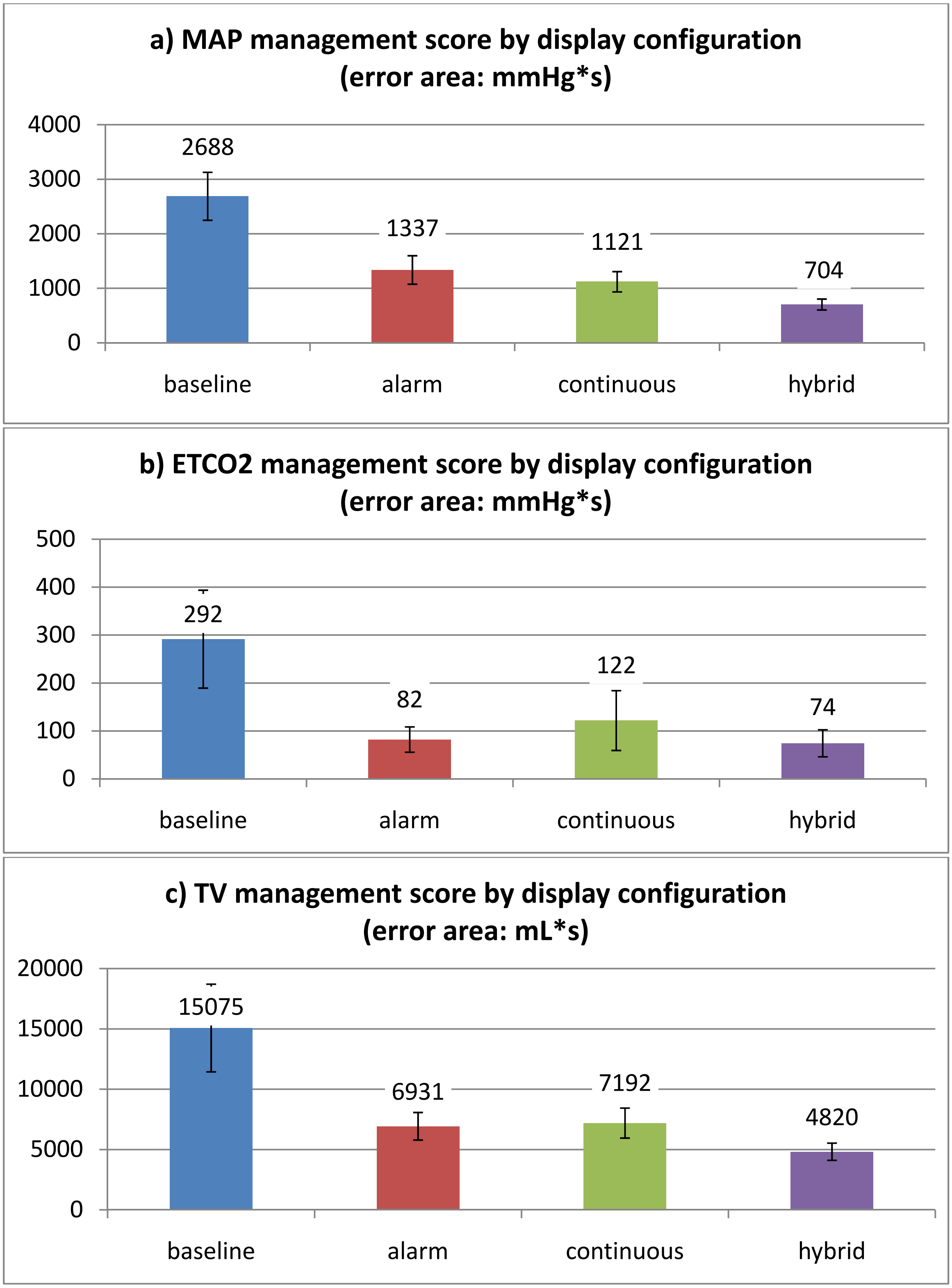


Figure 5.29: Physiological management score according to display configuration for a) MAP, b) ETCO2, and c) TV. Error bars represent standard error. Significant effects of display configuration were found for physiological management scores for MAP and TV, but not for ETCO2.

Discussion

H8: Participants will manage MAP, ETCO2, and TV levels more effectively when equipped with any of the three vibrotactile displays, when compared to the baseline display configuration.

In regard to physiological management of both MAP and TV, H8 is confirmed. For these parameters, a significant effect of display configuration shows that the baseline configuration resulted in the highest (worst) management scores, when compared to any of the tactile configurations. Although this effect did not reach significance for the ETCO2 parameter, the pattern in the data (see Figure 5.29b) suggests a trend similar to the effect found for MAP and TV, with the baseline configuration showing larger amounts of error area (292 mmHg*s) than any of the tactile configurations (more than twice the score of the second-largest score: 122 mmHg*s; and nearly four times as large as the smallest score: 74 mmHg*s).

One reason why the ETCO2 management scores were not significantly affected by display configuration, when the scores for the other parameters were, could be the relatively fewer opportunities for ETCO2 to reach problematic levels in this experiment. The end result is that ETCO2 error areas were not numerous or large enough to observe a significant effect of display configuration. Only one respiration event – hypercarbia – affected solely ETCO2; the other respiration and emergency events that could cause ETCO2 to reach levels outside of the acceptable range tended to do so only in the later stages of event development. These events were frequently detected and corrected before ETCO2 levels exceeded its acceptable range. Two exceptions to this rule were found in the shortshallow1 and shortshallow2 “nervous spontaneous breathing” events: in these cases, a high respiration rate led to ETCO2 levels exceeding the acceptable range before breaths were shallow enough to exceed the acceptable range for TV levels. It is worth noting here that for these two events, as well as for hypercarbia, the detection and correction times were significantly affected by display configuration, showing the (significantly) worst performance in the baseline configuration (see Figures 6.24a and 6.24b, Figures 6.25a and 6.25b, and Figures 6.26a and 6.26b). These findings suggest that the inclusion of more events that involve problematic ETCO2 levels as a primary characteristic may have led to a clearer reflection of the expected benefit of the tactile displays in ETCO2 management scores.

Other than for the three exceptions mentioned, the vast majority of events in this study showed significant changes in, or abnormal levels of MAP and/or TV before critical levels of ETCO2 could be realized. This reflects a causal relationship between these parameters which is relied upon in clinical practice. Unlike the parameters MAP and TV, which can change relatively independently in response to various events and actions, ETCO2 is strongly correlated with, and often dependent on levels/changes in MAP and/or TV. For example, if MAP reaches very low levels, there may not be sufficient pressures to support adequate blood perfusion of organs and muscles. This means CO2 waste will not be adequately removed from these systems by the blood and delivered to the lungs for elimination by exhalation, which is the process that defines ETCO2 levels. Following this relationship, events in this study which led to very low MAP levels – emergency events pneumothorax, anaphylaxis, and PEA – would eventually lead to very low ETCO2 levels as well. The fact that low ETCO2 levels were rarely and/or only briefly realized reflects that these events tended to be recognized before this point. Similarly, low or high TV levels are a major causative factor in high or low ETCO2 levels, respectively, but abnormalities in the ETCO2 level take longer to develop. Events in this study which caused abnormal TV levels also tended to be detected and corrected before problematic ETCO2 states were realized. The underlying message here is that for the vast majority of events in this study, participants may have (correctly) prioritized monitoring for changes and abnormal levels in MAP and TV over ETCO2. Thus the MAP and TV management scores can be considered more indicative of overall physiological management performance than the ETCO2 score, and the former two scores showed a significant improvement for each tactile configuration over the baseline configuration.

H9: Participants will manage these levels most effectively with the hybrid display.

H9 is partially confirmed, in the cases of MAP and TV management. In each case, the hybrid display significantly outperformed the baseline configuration, and showed at least a strong trend toward lower (better) management scores than the alarm and continuous configurations. Because the physiological management score was related to the ability to a) quickly detect and b) quickly correct events, this finding falls in line with the significant benefit that the hybrid display showed for faster detection and correction times (see Figure 5.21).

Related to this hypothesis, it is interesting to note that, across all participants, 9 of the 320 total events were detected during the “trend” phase of event development and corrected without ever realizing an exceedance of the boundary of the acceptable range for any physiological parameter. Eight of these 9 examples of the ideal response to health events were observed under the hybrid display configuration, representing 10% of all events that occurred under this configuration. One of the primary goals of a continuously informing display is to recognize the early signs of developing events and correct them *before* they significantly affect the health of the patient. This observation suggests the hybrid display may provide a substantial advantage in achieving that goal.

4) Induction Completion Time

Results

Figure 5.30 shows induction completion times for each display configuration. Neither *display configuration* nor *experience as a practicing anesthesiologist* significantly affected this measure. However, induction completion times were affected by scenario ($F(3,13) = 19.754$; $p < .001$) (see Figure 5.31). Post-hoc tests showed that Scenario 1 showed the longest completion times (mean: 897 s, $p < .001$ for all pairwise comparisons with other scenarios), and Scenario 2 (781 s) showed longer times than Scenario 3 (716 s; $p = .015$) and Scenario 4 (695 s; $p < .001$). Completion times for Scenario 3 and Scenario 4 did not differ significantly.

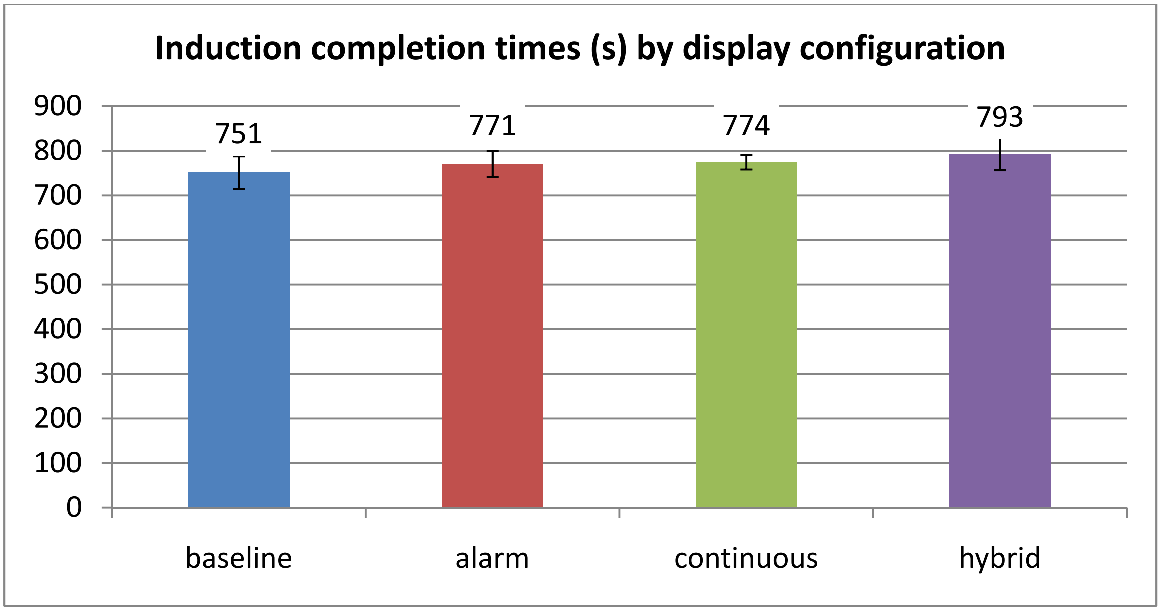


Figure 5.30: Induction completion times under each display configuration. Error bars represent standard error. No significant differences were found due to display configuration.

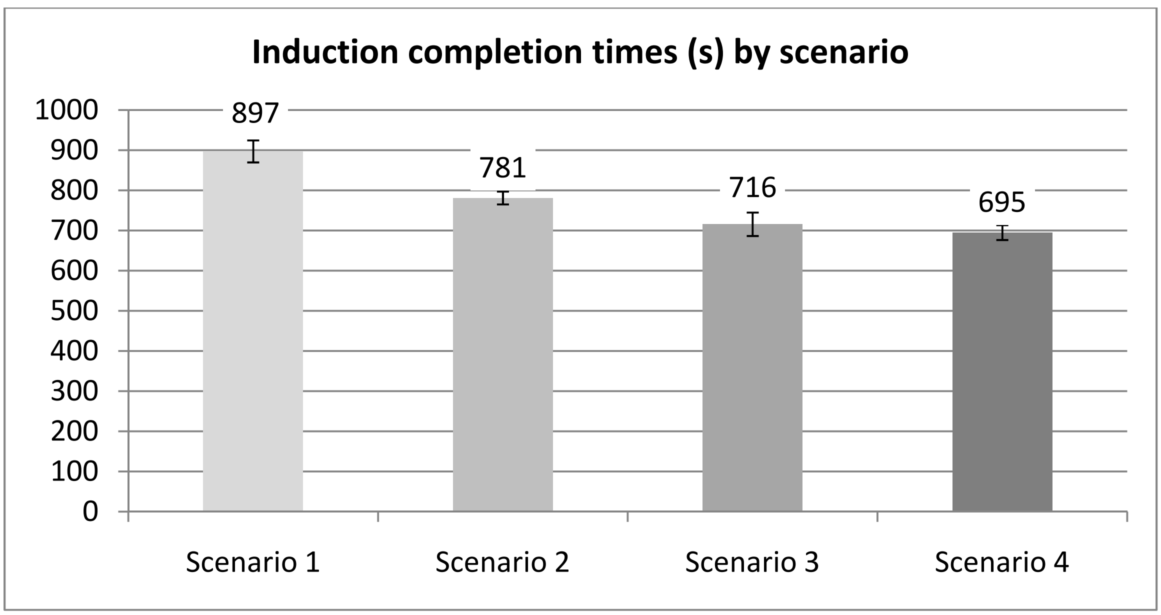


Figure 5.31: Induction completion times for each scenario. Error bars represent standard error. Later scenarios were completed significantly faster than earlier ones.

Discussion

In contrast to the findings for measures related to the physiological management task, a difference in induction completion time was found due to the factor *scenario*. This data is reported to highlight a potential limitation of this study. One possible concern regarding the experimental design was that a learning effect across scenarios might cause a higher degree of variance in performance measures during earlier scenarios. This higher performance variation may cause the effects of display configuration to wash out, at least in earlier scenarios. One way that this could have been resolved would be to include the *scenario* * *display configuration* interaction effect in the statistical models; however, it was deemed inappropriate to test for an interaction effect between scenario and display configuration, since this interaction was confounded by the between-subjects variable of *subject*. The expectation was that each subject would show a different level of baseline performance, hence the repeated measures statistical model was used.

The significant effect of scenario may have influenced the variance for induction completion times in the analysis of this measure across display configurations. If it were appropriate to test for an interaction effect, it may have been observed, for example, that the apparent (though slight) trend toward longer completion times in the hybrid display configuration could reach significance in some individual scenarios. However, with the experimental design employed in this study, the only conclusion that can be drawn is that induction completion times were not significantly affected by display configuration.

H10: The patient induction task will be completed faster when participants are equipped with any of the three vibrotactile displays, when compared to the baseline display configuration.
H11: The patient induction task steps will be completed fastest with the hybrid display.

Neither H10 nor H11 are confirmed, since no significant effect was found for display configuration on induction completion times. However, this lack of effect may not be an entirely negative result, since none of the tactile displays significantly hurt performance on the induction task. One observed aspect of participant behavior may explain the lack of benefit for tactile displays. In the baseline configuration, many participants appeared to recognize the disadvantage they faced in the physiological management task and therefore shifted their priorities to focus more completely on high performance in the induction task. This shift is reflected in the substantially worse performance under this configuration for all the measures associated with the physiological monitoring task, and the trend toward slightly faster induction completion times.

While this priority shift under the baseline configuration ultimately proved to be an unwise strategy for the overall goal of maximizing performance on both tasks in this study, it actually somewhat accurately reflects the task priority that anesthesiologists normally practice in the clinical setting. Several participants noted, either in conversation with the experimenter or explicitly in the postexperiment questionnaire (see Appendix 8), that their priorities under most normal conditions are to first assure that the patient's airway is secure and secondarily to assure that breathing and circulation are normal (commonly referred to as the ABC's: Airway, Breathing, and Circulation responsibilities). At times, this may mean postponing addressing problems related to breathing and circulation when the airway has not yet been secured. For the purposes of this study, participants were explicitly instructed to treat the task of securing the airway (steps of the induction task) with an equivalent priority as managing breathing and circulation (the physiological monitoring task). Possibly, under the baseline display configuration, participants may have reverted to a more natural task priority mapping than that which was instructed. This conflict describes a potential limitation of this study.

5) Multitask Performance Score

Results

Multitask performance scores for each participant in each scenario were calculated according to Equation 5.1. Figure 5.32 shows the means for each of these scores across participants for each display configuration. A repeated measures ANOVA found a significant effect of *display configuration* on multitask performance score ($F(3,13) = 9.623$; $p = .001$). *Experience as a practicing anesthesiologist* did not show a significant effect. Post-hoc tests found the baseline configuration (mean score: 1.451) showed significantly worse multitask performance than did the alarm (0.924; p = .001), continuous (0.875; p = .003), and hybrid (0.750; p < .001) configurations. Further, the hybrid configuration showed significantly better multitask performance than the alarm (p = .018) and continuous (p = .048) configurations. Multitask performance scores for the alarm and continuous configurations did not differ significantly.

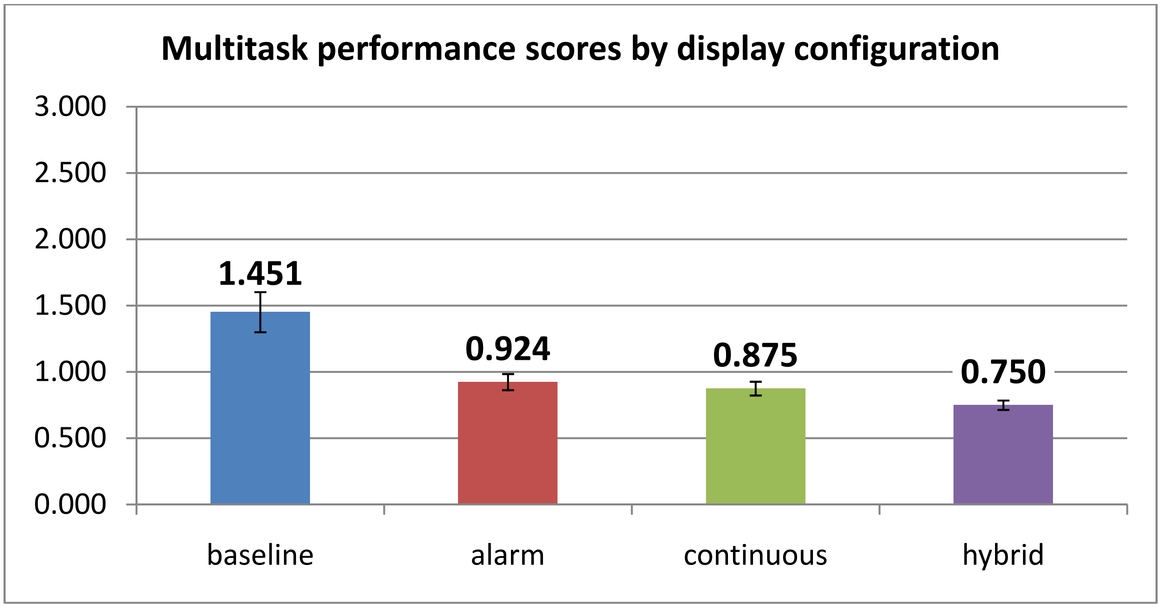


Figure 5.32: Mean multitask performance scores, calculated with Equation 5.1, for each display configuration. Error bars represent standard error. A score of "1" represents mean performance across all scenarios/display configurations, and higher scores represent worse multitask performance.

Discussion

> **H14:** Scores for combined performance on both the induction task and the physiological management task will be better when participants are equipped with any of the three vibrotactile displays, compared to the baseline display configuration.
>
> **H15**: Combined performance scores will be best with the hybrid display configuration.

Both H14 and H15 are confirmed: multitask performance improves significantly with any of the tactile displays when compared to the baseline configuration, and the hybrid display led to the best overall performance, significantly better than both the alarm and continuous configurations. The score values can be roughly interpreted as a percentage, with 100% (scores of 1.0) representing the overall mean multitasking performance across all participants, scenarios, and display configurations. With this interpretation, compared to performance in the baseline configuration, overall multitask performance can be considered to be roughly 53% improved with the alarm configuration, 58% improved in the continuous configuration, and 70% improved in the hybrid configuration.

Recall that half of the multitask performance score was determined by the relative performance on the induction task (see Equation 5.1), and that no difference was found for this factor due to display configuration. Additionally, the three components of Equation 5.1 that represented the physiological monitoring task were the management scores for MAP, TV, and ETCO2, the last of which was not significantly affected by display configuration. Therefore, it can be concluded that the large majority for the performance benefit of the tactile displays, especially the hybrid display, are due to the ability to support the management of MAP and TV.

Another way to interpret the fact that significant differences were found for overall multitasking performance score, but not for all of the performance measures considered as components to this score, is that participants tended to handle the task tradeoff with different emphases. Some participants may have emphasized the physiological monitoring task, others the induction task, and still others may have changed their emphases between scenarios/display configurations. These differences may have led to higher variance in the individual performance measures that contributed to the lack of a significant effect being identified. The value of the multitask performance score defined by Equation 5.1 is that it allows comparison of overall performance in a way that controls for the fact that the tasks may have been emphasized differently between participants/scenarios/configurations.

6) Subjective Workload Ratings (NASA-TLX Survey)

Results

Following the completion of each scenario, participants rated their experience on 20-point analog scales in regard to six attributes: *mental demand, physical demand,* and *temporal demand* imposed by the scenario, their subject evaluation of their *performance*, the *effort* required to conduct the required tasks, and the *frustration* they experienced while doing so. Figure 5.33 shows mean ratings along each attribute, according to display configuration. Repeated measures ANOVAs were used to analyze the ratings, however none of the measures were found to be significantly affected by *display configuration*. One notable trend was found for *mental demand* (Figure 5.33a: $F(3,13) = 2.608$; $p = .096$), which seemed to suggest high perceived mental demand for the hybrid display and low demand for the alarm display.

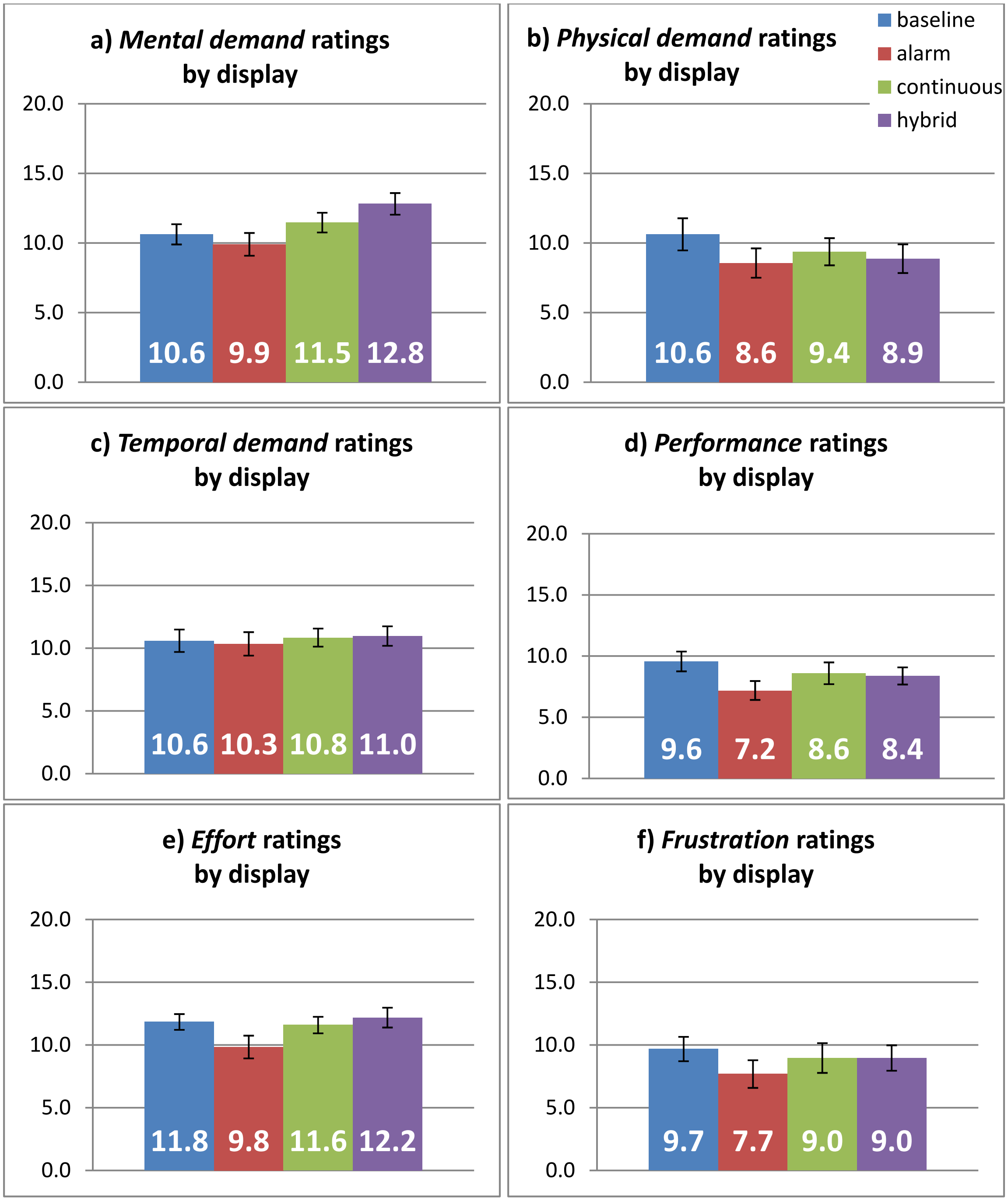


Figure 5.33: Mean ratings for each display configuration (key in upper right corner of (b)) for the six attributes on the NASA-TLX survey: a) Mental demand, b) Physical demand, c) Temporal demand, d) Performance, e) Effort, and f) Frustration. Refer to Appendix 7 for this survey, including exact descriptors of attributes. Error bars represent standard error. Generally, lower ratings represent more desirable traits; for *performance* ratings (d), lower ratings represent better perceived performance.

Discussion

H13: Participants will perceive and rate mental demand and effort (measured by NASA-TLX) to be the same for all display configurations.

H13 is confirmed, since display configuration did not show a significant effect on the ratings for any of the attributes. However, it is worth highlighting the trend found for mental demand, suggesting participants felt the demand was slightly higher in the case of the hybrid display. Participants were instructed to rate each attribute according to their overall experience conducting the multitask set in the given scenario, which may be affected by but should not have depended entirely on the display configuration. This means that participants may have felt higher levels of mental demand because the hybrid display imposed this demand, but another interpretation could be that the higher demand was imposed by the actions required to maintain a higher level of performance in the multitask set, and the hybrid displays were most often associated with these higher performance levels. This reasoning may also help explain why the ratings of effort (Figure 5.33e) seemed to show a similar (insignificant) pattern, for which the greatest effort was required in scenarios conducted under the hybrid configuration.

When reviewing these ratings, one other thing to keep in mind is that participants were not informed when they missed a serious health event. Since participants were exposed to each type of event in the training scenarios, they might have expected that they would encounter each type of event at some point in the experiment (though instructions were that the trained events *could* happen but wouldn't necessarily). Therefore participants were not informed when they missed an event, because doing so could have altered expectations of that event's occurrence in later scenarios in a way that adversely impacted performance. Because participants were oblivious of missed events, they could not consider them in any of these subjective ratings. Naturally, it could be expected that ratings for at least some attributes would be higher if the events were not missed, since higher levels of workload could be imposed by maintaining higher vigilance in the monitoring task and conducting additional corrective actions in order to address the events that were missed. Since the event detection rates for the baseline configuration were considerably (significantly) lower than with each of the three tactile displays, this expectation should be considered when interpreting the lower-than-expected subjective workload ratings for the baseline configuration.

7) Subjective Rankings of Display Configurations

Results

Figure 5.34 shows box plots of participants' final rankings of each display configuration in terms of a) *supporting multitasking performance*, b) *minimizing cognitive demand in multitask set*, c) *minimizing annoyance*, d) *minimizing unnecessary distraction*, e) *maximizing trust*, f) *comfort*, and g) *overall preference*. The rank data were analyzed with nonparametric Friedman tests. The results of these findings are found below Figure 5.34 in Table 5.7.

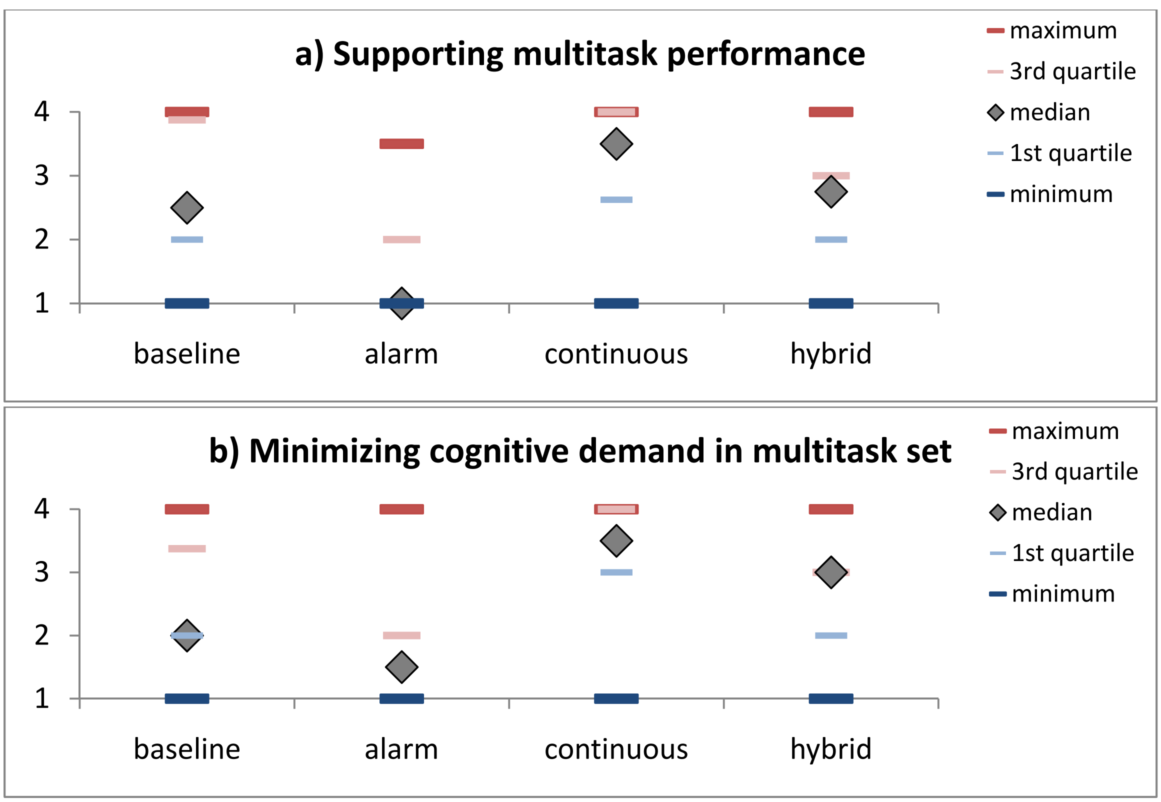


Figure 5.34: Box plots of subjective rankings of each display configuration (continued on next page).

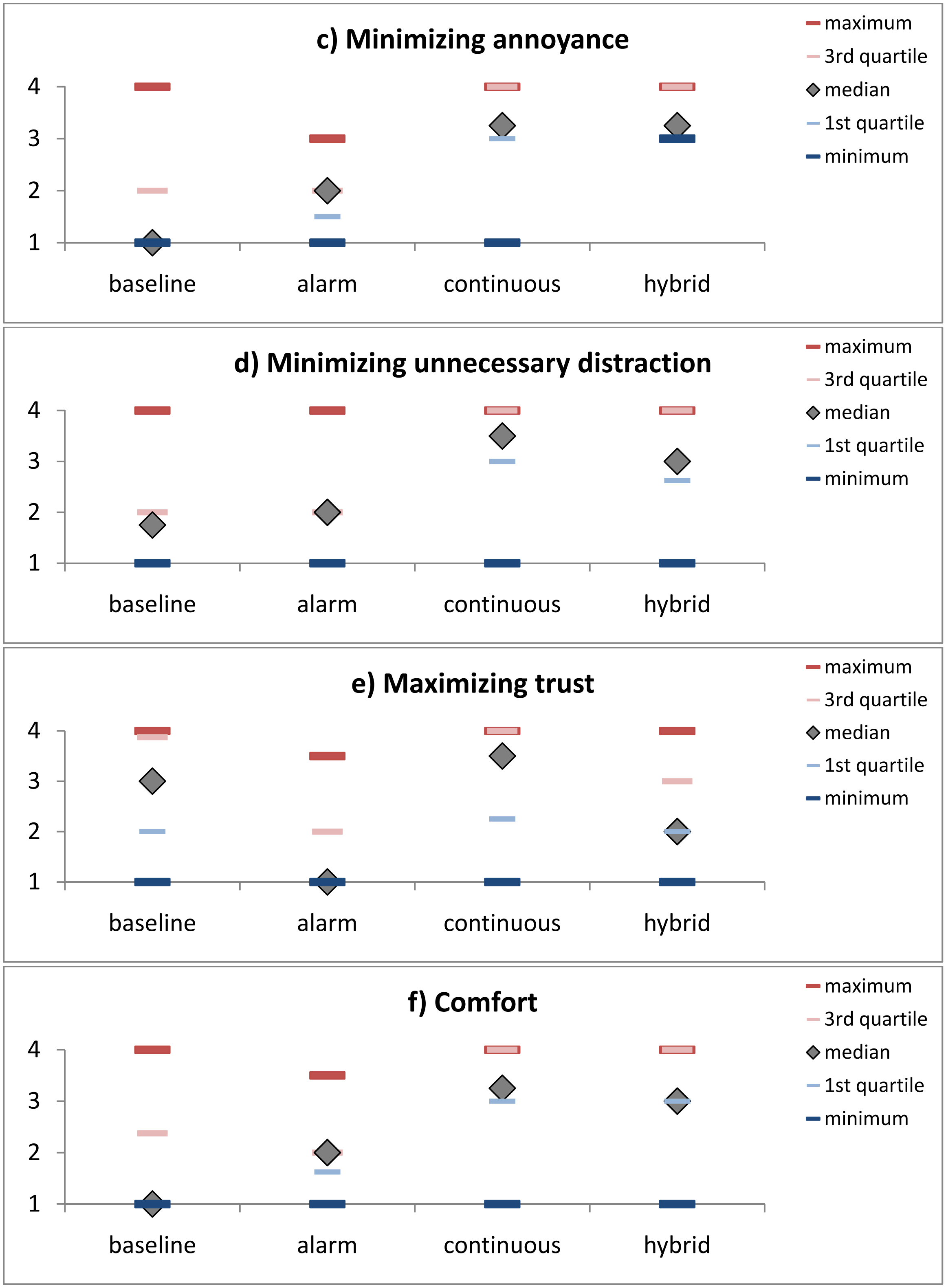


Figure 5.34: Box plots of subjective rankings of each display configuration (continued on next page).

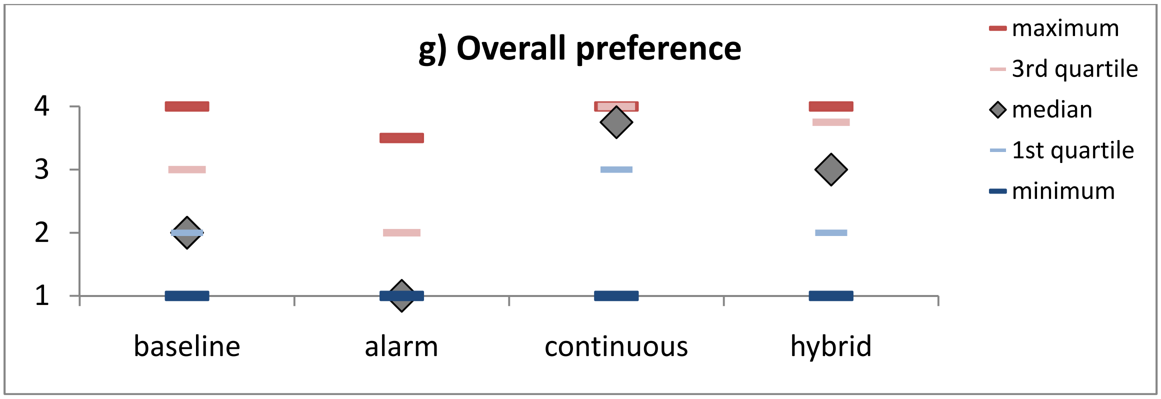


Figure 5.34: Box plots of subjective rankings of each display configuration, according to the attributes listed on the postexperiment questionnaire (see Appendix 8).

Attribute	Friedman test	Display configuration rank order (mean rank)
Supporting multitasking performance	$X^2(3, N = 16) = 14.477; p = .002$	alm(1.56) < hyb(2.56) < base(2.63) < cont(3.25)
Minimizing cognitive demand in multitask set	$X^2(3, N = 16) = 10.586; p = .014$	alm(1.78) < base(2.34) < hyb(2.66) < cont(3.22)
Minimizing annoyance	$X^2(3, N = 16) = 26.841; p < .001$	base(1.53) < alm(1.88) < cont(3.16) < hyb(3.44)
Minimizing unnecessary distraction	$X^2(3, N = 16) = 18.910; p < .001$	base(1.81) == alm(1.81) < hyb(3.09) < cont(3.28)
Maximizing trust	$X^2(3, N = 16) = 14.428; p = .002$	alm(1.59) < hyb(2.38) < base(2.84) < cont(3.19)
Comfort	$X^2(3, N = 16) = 21.122; p < .001$	base(1.66) < alm(1.94) < hyb(3.16) < cont(3.25)
Overall preference	$X^2(3, N = 16) = 17.791; p < .001$	alm(1.47) < base(2.41) < hyb(2.81) < cont(3.31)

Table 5.7: Significance levels and mean subjective rank order of display configurations for each of the attributes listed on the postexperiment questionnaire (see Appendix 8).

Discussion

To summarize the results of the subjective rankings, every attribute showed a significant difference among display configurations. Generally, participants liked the alarm display best and the continuous display the least. Participants tended to rank the baseline configuration as best and the alarm configuration second for attributes which did not strongly relate to performance (*minimizing annoyance* and *comfort*). This reflects a sentiment that it was not entirely pleasant to receive frequent vibration presentations. While the primary goals/questions of this study did not include determining which tactile display designs were most comfortable or least annoying, the rankings along these attributes help gauge overall attitudes about the displays, which may have directly impacted some of the subjective rankings in other attributes. For example, participants clearly favored the baseline configuration over the hybrid and continuous displays in their rankings of *overall preference*. Because performance so clearly benefitted when either tactile display was equipped, when compared to the baseline configuration, the amount of annoyance/discomfort each display caused the participants likely impacted their overall preference as much or more so than did their ability to support performance.

H16: Participants will subjectively rank the baseline display configuration as worst, and the hybrid display configuration as best, in terms of the ability to support multitask performance.

H16 is rejected. The ranking for the ability to support multitask performance instead strongly favored the alarm display. The hybrid display ranked second, followed by the baseline and continuous configurations. It was surprising to see such a mismatch between objective performance (which clearly favored the hybrid display) and subjective ranking. However, as was the case when interpreting the NASA-TLX ratings, it is important to remember that participants were often naïve to their own poor performance levels because they were not made aware of missed events, nor were they told how long detected events were active before the participants detected them. In retrospect, the expectation that drove the formulation of this hypothesis – that participants would be able to recognize when the tactile displays best supported multitasking – is likely not valid, and this is reflected in the ranking for this attribute.

It is telling that participants felt that their multitasking performance was best supported with the alarm display. They also rated the alarm display highest in terms of minimizing cognitive demand (which also reflects some trends in the NASA-TLX ratings), and maximizing trust in the signal to attract attention when it was warranted. Despite keeping an anesthesiologist less informed about the physiological parameters of interest, the alarm display's more simplified presentation may be more universally acceptable for use in a clinical setting.

H17: Participants' subjective rankings of the alarm configuration's ability to minimize unnecessary distraction will be worse than for the continuous or hybrid display configurations.

H17 is rejected, as results showed the opposite of the expected pattern. It is not surprising that the baseline configuration was ranked as best supporting this attribute, because this configuration minimized *all* exogenous attention orientation, necessary and unnecessary. Between the three tactile configurations, the alarm display was ranked as best for this attribute, followed by the hybrid and continuous displays. This ordering likely reflects the fact that some brief trends in physiological parameter levels – which were communicated by the continuous and hybrid displays but not the alarm display – often did not lead to serious health events. Participants may have classified the notifications associated with these trends as unnecessary distractions when serious events did not follow. Interestingly, participants did not apply this same logic to the alarm display, which showed nearly equivalent rankings with the baseline configuration. Each scenario featured "benign" events that would activate the tactile alarms, for example, low TV alarms when the patient became apneic following propofol administration and high ETCO2 alarms whenever a respiration was registered following a prolonged apneic period. Perhaps because these alarms were investigated and quickly dismissed, participants did not judge them to be overly distracting. Given anesthesiologists' familiarity with these types of "nuisance" alarms (though announced with auditory signals), a practice effect may also be present, in that once participants became accustomed to the tactile signals, they were able to recognize and dismiss them as easily as their auditory counterparts.

It is worth noting that participants seemed to appreciate the "beat" pulse that the hybrid display utilized to denote parameter levels passing graded thresholds. This ability to more clearly distinguish parameter levels and judge their severity likely reflects the slight ranking benefit the hybrid display showed over the continuous display in regard to minimizing distraction.

Finally, as with the other performance-related attributes, the unexpected significant ranking pattern of the display configurations for this attribute may somewhat reflect the participants' attitudes about vibration displays in regard to comfort and annoyance. Discussions with some participants following the experiment, and others' responses to questions on the postexperiment questionnaire (Appendix 8), seemed to suggest that these attitudes could not easily be set aside when participants judged the displays for their abilities to support performance. While the word "distraction" in the context of the study was intended to mean "something that captures attention when it ought not to, thus hurting performance on tasks where attention is required", it apparently was often interpreted by participants as "something that is annoying." While these interpretations are related, the latter does not precisely address the goals of the study. Therefore, greater emphasis should be given to the objective performance measures than subjective rankings and ratings of the displays.

GENERAL DISCUSSION

This chapter described an empirical evaluation of three vibrotactile displays (alarm, continuous, and hybrid) which were designed to support aspects of "preattentive reference" (Woods, 1995), and hence support (anesthesiologists') attention and task management, in the context of a simulation that was representative of their responsibilities in the clinical setting. Two of these displays, the "continuous" and "hybrid" displays, represent a new type of tactile display which has been described as a "tactification": a continuously-informing display that transforms traditionally visually-displayed physiological data into coded vibration patterns that are presented to the skin in realtime. The findings from the evaluation study show promise for tactile and tactification displays to support multitasking performance when ongoing tasks place high demand on visual and auditory resources. Overall, the best objective performance was found with the hybrid tactification display, which mapped the severity of the state of monitored variables to the salience of the signal and used graded notifications to more clearly communicate state changes. This display significantly outperformed all other display configurations in a measure of multitasking performance (*multitasking performance score*), which equally weighted performance in a monitoring task and a complex ongoing visual/auditory task set. While subjective rankings of the displays suggested a preference for the alarm display (and, in some cases, the baseline/visual-only configuration) over the tactification displays, ratings of attributes related to subjective workload measured immediately after each experimental scenario did not differ between the display configurations, and any supposed increase in workload did not significantly impact performance.

The following sections will discuss the above findings in more detail and relate them to aspects of preattentive reference and earlier studies on notification design. The first important prerequisite for supporting preattentive reference is that *display signals can be reliably perceived and interpreted while minimally interfering with processing required for concurrent tasks*. With this goal in mind, the tactile channel was chosen for the display of patient data so that interference could be minimized at the perceptual processing stage, since ongoing tasks primarily engaged the visual and auditory sensory channels (e.g., Wickens, 2002; 2008). Additional steps were taken to minimize perceptual interference between separate components of the tactification display, due to the effects of vibrotactile adaptation and vibrotactile masking. For example, the interval between the onsets of MAP presentations, set at 3 seconds, was long enough to minimize adaptation effects without significantly pushing the limits of tactile short-term memory. This interval also reduced masking effects by avoiding the maximum overlap between the respiration and MAP display components. Other interferences that could take effect at the early stages of processing could be attributed to forms of tactile change blindness. Chapter 2: Part A demonstrated how these effects could affect the detection of changes in vibration intensity, which was an important dimension for supporting natural mapping in the tactification displays. Design guidelines drawn from the findings of this study included using large discrete steps for intensity changes, rather than analog changes in intensity. This study also served to demonstrate the likelihood that a complex tactile signal would lead to greater difficulty in detecting intensity changes due to these effects, thus intensity was employed as only one of multiple redundant ways to communicate critical health information.

Interference can also take place at the cognitive stage of processing, “downstream” of perception (Wickens, 2002; 2008), and Chapter 2: Part B detailed how the methods used to encode information into a vibrotactile pattern can have a significant effect on the degree of this interference, depending on the types of processing codes required for ongoing tasks. In order to minimize interference at the cognitive stage, the tactification displays were designed to be interpretable by engaging either of the two processing codes: spatial or nonspatial/symbolic processing. For example, MAP levels could be inferred either by the spatial location of the vibration presentation, or the intensity of the vibration. This redundancy in encoding method allowed the critical information to be extracted while engaging the processing code which faced less competition by concurrent (visual and auditory) tasks, hence minimizing interference at this stage.

The findings from the evaluation study suggest that all three tactile displays (the alarm, continuous, and hybrid configurations) successfully supported this first aspect of preattentive reference. Detection rates for serious health events were significantly higher under each tactile display configuration when compared to the baseline (visual-only) configuration. Further, while the difference did not reach significance, not a single event was missed under the hybrid configuration, but approximately 3% of events in both the alarm and continuous configurations were, possibly suggesting a higher tendency for perceptual and/or cognitive interference. The fact that detection times were fastest overall with the hybrid display suggests that its signal properties – which include the ability to support "continuously informing" but also to clearly communicate state changes through its "beat pulse" graded notifications – made it the most resilient to interferences that might delay detection under the other configurations. Perceptual and cognitive interferences might also be manifested as performance costs in ongoing tasks. The fact that induction completion times did not significantly differ among display configurations suggests that interpreting (visual and) tactile displays did not interfere with the induction task any more so than did interpreting (solely) the visual displays in the baseline configuration.

A second important prerequisite for supporting preattentive reference is that *signals include partial information on what attention-directing signals refer to, so the operator can infer whether the signal warrants a shift in attention or not.* The survey described in Chapter 3 helped identify two general categories of information as important to the decision of whether or not a physiological parameter warrants a shift in attention away from an ongoing task: the approximate level of the parameter and its rate of change. Following this survey, the study described in Chapter 4 showed how iconic tactile displays that relate level information were more useful in supporting this decision than those relating rate information. Displays that included *both* the level and rate information did not show a significant advantage over presenting solely the level information, despite the fact that full knowledge of both types of information should be expected to improve decisions about when to switch attention (for example, avoiding a switch when levels are high, but trending downward). It was expected that the lack of benefit for rate information was due to difficulty in interpreting the rate presentations, possibly because this information was presented in a somewhat unnatural way:

abstracting the amount of change over a set time window into an iconic pulse pattern which was considerably shorter than the represented time window.

With regard to relating the rate of change in a parameter, the tactification displays represent an improvement over the iconic displays in Chapter 4. For the tactifications, the rate information is not abstracted; instead, the natural progression of the continuous signal allows rate of change in a monitored parameter to be inferred by noting the frequency of level changes in realtime. The continuous and hybrid displays related parameter levels as within 7 different level ranges (3 level thresholds on either side of the “normal” range), which allowed a reasonable resolution sufficient for distinguishing “slow” trends from “fast” trends.

The hybrid display showed improved performance over the alarm display in the physiological monitoring task, in terms of faster detection and correction times and better management scores. Since the hybrid display supported the ability to infer rates of change while the alarm display did not, this suggests that access to the rate information can lead to better decision making about when to correctly switch attention from the induction task to the physiological monitoring task. It is harder to gauge from the experimental data how the two displays compared in terms of supporting when to correctly *not* switch attention away from the induction task (e.g., when no serious health event is present). However, since no significant difference was found between the two display configurations in terms of induction completion times, it can be assumed that participants did not inappropriately switch their attention more often (or for longer periods) with the hybrid display than with the alarm display.

Though it was not captured by any experimental measures, an interesting observed behavior suggests an additional benefit for the tactification displays over the alarm display. Each scenario involved at least two “nuisance alarm” events, for which normal and expected occurrences led to parameter levels that technically constituted “problem” states, in the sense that high/low levels caused an activation of alarms, not in the sense that any problem needed to be corrected. For example, after administration of propofol, the patient’s breaths became gradually shallower, eventually leading to apnea. When participants were equipped with the alarm display, they would often be startled and caught off-guard when the tactile alarm activated to signal low TV levels (and persisted while the patient was still breathing; alarms ceased when the patient reached an apneic state). Frequently these alarms led to disruption of the induction task to investigate the issue, returning to the task only when they remembered/realized/were reminded that shallow breathing leading to apnea was one of the intended effects of propofol administration. In contrast, when participants equipped with the continuous and hybrid displays encountered this same situation, they were able to feel the progression of shallower breaths, which began shortly after they administered propofol. Because they were much more aware of the situation as it developed, participants were much less likely to interrupt the induction task even when the displays communicated that TV levels fell within a "problem" state.

Interestingly, the continuous display did not show the same benefit over the alarm display that the hybrid display did, though the two tactifications relayed level and rate data for each parameter with the same resolution. This might reflect differences in the signal-mapping rules employed by the two tactifications. While the continuous display mapped all the dimensions of its signal to the absolute levels of parameters, the hybrid display used some signal dimensions to communicate the severity of the situation, as defined by the distance from a "normal" state. Since ultimately the participants needed to judge the situation's seriousness/severity when deciding whether to switch attention to the physiological monitoring task, the ability to infer this information directly from the hybrid signal (rather than calculating severity as a function of parameter level) may have led to an advantage. A note of caution should be given here, however, since in this case the operating definition of situation "severity" – the distance a parameter level was outside of "normal" states – will not always completely define the seriousness of a situation (e.g., in cases of the "nuisance alarm" events). One must take care to not become overreliant on a context-insensitive system to communicate the severity of the situation and dictate when to shift attention. Nevertheless, for the current study, this mapping provided the best support for decision making regarding when to switch attention.

Finally, *preattentive reference calls for signals that can be assessed in a mentally economical way.* One of the best ways to facilitate interpretation of the tactile signals – to support their assessment in a mentally economical way – is to exploit natural mapping between the signal modulation and the represented data (Norman, 2002). All three tactile displays exploited the spatial discrimination ability of the sense of touch to naturally map each represented parameter to body location, and the level (high or low) and direction of change (up or down) in those parameters to spatial direction (i.e., up and down the arm or back). The tactifications employed metaphorically-derived tactile icons to further help distinguish vibration presentations by providing an additional means of naturally mapping them to their respective parameters. Redundant encoding methods for each parameter level also exploited natural mappings. For example, increasing vibration intensity was used to relay increasing blood pressure for the MAP display component and increasing pressure in the lungs (i.e., higher TV levels) for the TV component.

One important dynamic of the pulse oximetry display – the "gold standard" for preattentive reference displays – which the tactification designs strove to emulate was its ability to exist in the attentional periphery, or "ambience" under normal conditions. Such a signal is always present and available for sampling with minimal effort, supporting reassurance without unduly capturing attention. This represents the highest form of supporting assessment in a mentally economical way. Then as a serious health situation develops, the natural progression of the pulse oximetry signal leads to a transition out of the periphery/ambience and into focal attention. The more dramatic the *change* in the signal, the more salient it is and thus the stronger the tendency for this transition to occur is. The two tactification displays used different techniques to achieve this dynamic. Both employed the same (relatively) subtle display patterns when parameters were within a "normal" state/range, designed to exist in the attentional periphery. While the continuous display was designed to closely follow the model of the pulse oximetry display – allowing changes in its signal to naturally support a transition to a higher attentional state when one was necessary – the hybrid display used more direct means of supporting this transition, by mapping the salience of the signal to the severity of parameter levels. Thus, while both displays relied on change in the signal to serve as a driver for this attentional transition, the hybrid display additionally triggered the transition when levels were sufficiently "severe". For the purposes of the evaluation study, the method employed by the hybrid display was more successful at supporting assessment in a mentally economical way: it not only led to more reliable and faster recognition of problem states, but was generally ranked higher (better) than the continuous display in subjective measures of *minimizing cognitive demand, minimizing unnecessary distraction,* and *maximizing trust in the signal* [to reliably communicate a "problem" state].

Of the display configurations evaluated in this study, it can generally be said that the hybrid display stood out as best supporting the three properties that define preattentive reference displays when the properties were considered altogether. It cannot be said, however, that the hybrid display is on par with the variable-tone pulse oximetry display in terms of its ability to support the overall concept of preattentive reference. This is clear from participants' subjective rankings of the displays, which suggested that the two tactification displays imposed more cognitive demand and led to more distraction than the alarm displays. While some of these rankings may be strongly influenced by participants' comfort levels and attitudes about being subjected to vibration presentations over extended periods of time, these are not entirely surprising findings. In responses to the post-experiment questionnaire and in casual conversations, some participants reported sentiments along the lines of the tactification displays being difficult to follow, with three going as far as to describe a feeling of "sensory overload" at times. Many expressed a desire for longer training times (which were on the order of 20 to 30 minutes) with the displays, and this may have indeed improved the subjective experience with the tactification displays.

Overall, the most preferred display configuration was the alarm display, with 11 of the 16 participants ranking it this way according to the postexperiment questionnaire. Two participants ranked the hybrid display, and one ranked the continuous display as their overall preferred display. Two participants unexpectedly ranked the baseline configuration as the most preferred configuration. When considering these findings, it is important to note that participants were not aware of their overall performance under any configurations (e.g., they were not told when they missed serious health events), which helps explain some of the discrepancies between subjective rankings and objective performance. For example, participants rated the alarm configuration highest, and the hybrid configuration as nearly equivalent to (though slightly better than) the baseline configuration for the attribute *supporting multitask performance* (see Figure 5.34a and Table 5.7). This pattern of rankings contrasts strongly with the objective multitasking measure – the multitask performance score – which showed the hybrid display supported significantly better performance, and the baseline configuration led to significantly worse performance, compared to all other configurations (see Figure 5.32).

This contrasting relationship between objective performance measures and subjective preferences is a not an uncommon finding when introducing new tools and technologies to humans in the work environment. People have a strong tendency to prefer interfaces/technologies/methods which they are already familiar with, even when they hinder performance (Andre & Wickens, 1995; Bailey, 1993; Kweon, Schlegel, & Purswell, 1992). This helps explain why some participants preferred baseline display configurations and most preferred the tactile alarms: these configurations most closely resembled those that anesthesiologists are already familiar with in the clinical setting.

It is interesting to note that preferences/attitudes regarding the tactification displays were not universal. Three participants ranked the two tactifications as the best two display configurations in terms of *supporting multitasking performance*, *minimizing cognitive demand*, *maximizing trust*, and *overall preference*. The two participants who ranked the hybrid display as best according to these 4 attributes also happened to be the ones who showed the best (lowest) multitasking performance scores averaged across all scenarios/configurations. A similar pattern of results was found in the study described in Chapter 4, which presented blood pressure data via iconic tactile displays of different degrees of informativeness/complexity. In this study, some participants thrived with the level+rate display, showing large performance benefits under that configuration compared to the other simpler but less informative displays. Others reported being “overwhelmed” at times by this most complex display, to the point that they tried to “tune out” the vibrations and rely on solely visual means to monitor blood pressure.

Some participants in both the current study and the one described in Chapter 4 reported experiencing “sensory overload” when attempting to process visual, auditory, *and* (continuous) tactile information concurrently. These reports, taken together with the empirical findings of the two studies, have implications for the development of models of the structure of mental resources, such as Multiple Resource Theory (e.g., Wickens, 2002; 2008). These models are built on solid foundations of empirical evidence regarding the concurrent processing of visual and auditory information, while relatively little evidence exists for multimodal information processing that involves the tactile channel. Even less is known about the ability to concurrently process task-relevant data via three (e.g., visual, auditory, and tactile) or more sensory modalities. In particular, the reports of “sensory overload” suggest that information processing that engages more than two sensory modalities could have an effect of lowering the “red-line” of cognitive workload (e.g., Grier et al., 2008; Wickens, 2008), showing decrements in processing ability that are a function of the number of engaged channels. This idea stands as a direction for future research.

The differences among participants with regard to preferences and performance levels under each display configuration suggest another idea for future research: adaptive and/or adaptable tactification designs. Adaptive tactifications, i.e., those that automatically change configuration as a function of situational context (e.g., Sarter, 2007), could support the anesthesiologist (or operators in other domains) by presenting relatively less information in a simpler format during times when demand for processing resources are high. Recent work has shown some promise for adaptive display algorithms applied to auditory alarms in the clinical setting. For example, systems that automatically delayed the presentation of an alarm while a loud suctioning tool was in use showed substantial reductions in the number of ineffective and ignored alarms (Görges, Markewitz, & Westenskow, 2009). Similar methods could be used by integrated anesthesia systems to infer demand levels, and adjust the complexity of the tactification signal accordingly.

Adaptable designs, i.e., those that allow the anesthesiologist/operator to specify the format of the signal, could help account for individual differences in the ability to process the tactifications of varying complexity. “Scalable” tactifications could allow operators to determine how frequently vibrotactile updates are given for the levels of specified parameters, an idea that has been applied to sonification design (Watson, 2006). Another option could be to allow anesthesiologists the ability to choose which presentation modality, or modalities, to engage in the monitoring task. Auditory displays could then be reserved for the parameters that, like the pulse oximetry display, are most relevant for all clinical staff, while tactile displays can be utilized to communicate, in a private manner, data that is primarily of interest to the anesthesiologist.

REFERENCES

Andre, A.D. & Wickens, C.D. (1995). When users want what’s not best for them. *Ergonomics in Design, 3*(4), 10-14.

Bailey, R.W. (1993). Performance vs. preference. *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting*, 225-229.

Bowers, R.L., Mollenhauer, M.S., & Luxford, J. (1990). Short-term memory for tactile and temporal stimuli in a shared-attention recall task. *Perceptual and Motor Skills, 70*, 903-913.

Cholewiak, R.W. & Craig, J.C. (1984). Vibrotactile pattern recognition and discrimination at several body sites. *Perception & Psychophysics, 35*, 503-514.

Craig, J.C. (1983). Some factors affecting tactile pattern recognition. *International Journal of Neuroscience, 19*, 47-58.

Fitch, W.T. & Kramer, G. (1994). Sonifying the body electric: Superiority of an auditory over a visual display in a complex, multivariate system. In Kramer, G. (Ed.), *Auditory Display: Sonification, Audification, and Auditory Interfaces*. New York: Addison-Wesley.

Gaba, D.M. & Lee, T. (1990). Measuring the workload of the anesthesiologist. *Anesthesia Analgesia, 71*, 354-361.

Gallace, A., Tan, H.Z., Haggard, P., & Spence, C. (2008). Short term memory for tactile stimuli. *Brain Research, 1190*, 132-142.

Gallace, A., Tan, H.Z., & Spence, C. (2006). The failure to detect tactile change: A tactile analogue of visual change blindness. *Psychonomic Bulletin & Review, 13*(2), 300-303.

Gallace, A., Tan, H.Z., & Spence, C. (2007). Do “mudsplashes” induce tactile change blindness? *Perception & Psychophysics, 69*(4), 477-486.

Geldard, F.A. (1960). Some neglected possibilities of communication. *Science, 131*, 1583-1588.

Görges, M. & Staggers, N. (2007). Evaluations of physiological monitoring displays: A systematic review. *Journal of Clinical Monitoring and Computing, 22*, 45-66.

Görges, M., Markewitz, B.A., & Westenskow, D.R. (2009). Improving alarm performance in the medical intensive care unit using delays and clinical context. *Anesthesia & Analgesia, 108*, 1546-1552.

Greenspan, J.D. & Bolanowski, S.J. (1996). Psychophysics of tactile perception. In Kruger, L. (Ed.), *Pain and Touch*. San Diego: Academic Press, 25-103.

Grier, R., Wickens, C., Kaber, D., Strayer, D., Boehm-Davis, D., Trafton, J.G., & St. John, M. (2008). The red-line of workload: Theory, research, and design. *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting*, 1204-1208.

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

James, R.H. (2003). 1000 anesthetic incidents: Experience to date. *Anaesthesia, 58*, 856-863.

Kweon, O., Schlegel, R.E., & Purswell, J.L. (1992). Power window control location stereotypes. *Proceedings of the Human Factors and Ergonomics Society 36th Annual Meeting*, 485-489.

Liu, D., Görges, M., & Jenkins, S.A. (2010). Collection of vital signs data for anesthesia monitoring research. Abstract for the *International Anesthesia Research Society 2010 Annual Meeting*. Honolulu, HI: 20-22 March.

Loeb, R.G. (1993). A measure of intraoperative attention to monitor displays. *Anesthesia Analgesia, 76*, 337-341.

Loeb, R.G. (1994). Monitor surveillance and vigilance of anesthesia residents. *Anesthesiology, 80*, 527-533.

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

MacLean, K.E. (2008a). Foundations of transparency in tactile information design. *IEEE Transactions on Haptics, 1*(2), 84-95.

MacLean, K.E. (2008b). Haptic interaction design for everyday interfaces. In Carswell, M. (Ed.) *Reviews of Human Factors and Ergonomics, Vol. 4.* Santa Monica, CA: Human Factors and Ergonomics Society. 149-194.

Ng, J.Y.C. & Man, J.C.F. (2004). Vibro-monitor: A vibrotactile display for physiological data monitoring. *Human Interface Technologies Conference, University of British Columbia*, 1-8.

Norman, D.A. (2002). *The Design of Everyday Things*. New York: Basic Books.

Pasquero, J. (2006). Survey on communication through touch. Technical Report: TR-CIM 06.04. Center for Intelligent Machines, Department of Electrical and Computer Engineering, McGill University. August.

Rady, M.Y., Ryan, T., and Starr, N.J. (2001; 1998). Perioperative determinants of morbidity and mortality in elderly patients undergoing cardiac surgery. *Critical Care Medicine, 29*(9), S163-S172.

Sanderson, P.M., Liu, D., & Jenkins, S.A. (2009). Auditory displays in anesthesiology. *Current Opinion in Anesthesiology, 22*, 788-795.

Sanderson, P.M., Watson, M.O., & Russell, W. (2005). Advanced patient monitoring displays: Tools for continuous informing. *Anesthesia Analgesia, 101*(1): 161-168.

Sanderson, P.M., Watson, M.O., Russell, W.J., Jenkins, S., Liu, D., Green, N., Llewelyn, K. , Cole, P., Shek, V., & Krupenia, S.S. (2008). Advanced auditory displays and head-mounted displays: Advantages and disadvantages for monitoring by the distracted anesthesiologist. *Anesthesia Analgesia, 106*, 1787-1797.

Sarter, N. (2007). Coping with complexity through adaptive interface design. In Jacko, J. (Ed.), *Human-Computer Interaction, Lecture Notes in Computer Science, vol. 4552*, 493-498.

Seagull, F.J., Wickens, C., & Loeb, R.G. (2001). When is less more? Attention and workload in auditory, visual, and redundant patient-monitoring conditions. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting*, Santa Monica, CA, 1395-1399.

Trafton, J.G., & Monk, C.A. (2008). Task interruptions. In Boehm-Davis, D.A. (Ed.), *Reviews of human factors and ergonomics, Vol. 3*. Santa Monica, CA: Human Factors and Ergonomics Society. 111-126.

Verillo, R.T. & Gescheider, G.A. (1977). Effect of prior stimulation on vibrotactile thresholds. *Sensory Processes, 1*, 292-300.

Watson, M. (2006). Scalable earcons: Bridging the gap between intermittent and continuous auditory display. *Proceedings of the 12th International Conference on Auditory Display*, 59-62.

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

Webb, R.K., van der Walt, J.H., Runciman, W.B., Williamson, J.A., Cockings, J., Russell, W.J., and Helps, S. (1993). Which monitor? An analysis of 2000 incident reports. *Anaesthesia and Intensive Care, 21*, 529-542.

Welch, R.B. & Warren, D.H. (1986). Intersensory interactions. In Boff, K., Kaufman, L., & Thomas, J. (Eds.), *Handbook of Perception and Human Performance, Vol 1: Sensory Processes and Perception*. John Wiley & Sons.

Wickens, C.D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Sciences, 3*, 159-177.

Wickens, C.D. (2008). Multiple resources and mental workload. *Human Factors, 50*(3), 449-455.

Woods, D.D. (1995). The alarm problem and directed attention in dynamic fault management. *Ergonomics*, 38 (11), 2371-2393.

Chapter 6

Conclusion

The task set of an anesthesiologist, like that of operators in many other complex, data-rich domains, is characterized by high mental workload and the need for effective attention management. The inefficient allocation of anesthesiologists' attentional resources among sources of patient health data has been linked to monitoring errors, which constitute a significant portion of the preventable medical errors that are prevalent in current healthcare systems (Cooper, Newbower, Long, & McPeek, 1978; Kohn, Corrigan, & Donaldson, 2000; Walsh & Beatty, 2002; Webb et al., 1993). Researchers have sought to better support anesthesiologists' attention management, and hence reduce monitoring errors, through the design of improved alarm systems and advanced physiological display technologies. One promising approach is the introduction of non-visual "continuously informing" displays of patient health data. To date, audition is the one non-visual modality that has been employed for this purpose in the form of *sonifications*, i.e., the mapping of numerical values or relations in data to values or relations in one or more dimensions of an auditory signal (Kramer, 1994). With the variable-tone pulse oximetry sonification display serving as the "gold standard", complex sonifications have been designed that incorporate blood pressure and respiratory measures into the auditory signal (e.g., Fitch & Kramer, 1994; Loeb & Fitch, 2002; Sanderson et al., 2008; Seagull, Wickens, & Loeb, 2001; Watson & Sanderson, 2004). By supporting heightened awareness of these data, sonifications can improve monitoring performance and also performance on concurrent tasks, because they lessen the need to rely on context-insensitive auditory alarm systems. These systems are characterized by high rates of false and nuisance alarms (e.g., Imhoff & Kuhls, 2006; Seagull & Sanderson, 2001), and thus relying on them to direct attention tends to lead to frequent unnecessary distractions and interruptions of concurrent tasks, at the cost of overall multitasking performance (e.g., Trafton & Monk, 2008).

While complex sonification displays have shown some success in supporting monitoring and multitasking performance, they have not achieved levels of success on par with the pulse oximetry display. One reason for this may be that the more complex displays do not support the properties of *preattentive reference* as well as the pulse oximetry signal does. Displays that support preattentive reference: 1) are capable of being processed in parallel without interfering with ongoing tasks, 2) include partial information on the state of the displayed system, so the operator can infer whether and when a shift in attention is warranted to address the system, and 3) can be assessed in a mentally economical way (Woods, 1995). Complex sonifications are particularly challenged to meet the first property – supporting processing in parallel without interfering with ongoing tasks – as the auditory channel is becoming saturated with auditory signals and noise in the Operating Room (OR). These noise levels, which can come from surgical tools and equipment, necessary (and unnecessary) conversations between OR staff, and most notably, frequent false and nuisance auditory alarms, have been measured as comparable to sound levels when standing next to a busy highway (Hodge & Thompson, 1990; Ulrich et al., 2008).

To address the need to support more continuous awareness of key physiological parameters in an environment that imposes a heavy load on both the visual *and* auditory channels, a different approach was taken in the present research effort. Three types of vibrotactile displays – an alarm, a continuous, and a hybrid display – were developed and evaluated comparatively. In order to ensure the effectiveness of this approach, two empirical studies were conducted first to address the question of how to minimize interference a) within the tactile channel (i.e., between components of complex tactile displays), and b) at cognitive processing stages "downstream" of tactile perception. Chapter 2: Part A describes a study investigating the effects of a tactile analog of visual "change blindness" (i.e., within-channel interference) for the dimension of vibration intensity. This dimension was examined because it was identified, via the survey described in Chapter 3, as one that mapped most naturally to some of the information that was going to be displayed via tactification later on. It was critical to identify presentation patterns that were relatively resilient to these effects, thus minimizing perceptual interference between tactification display components. Chapter 2: Part B described a study that investigated potential interference between the interpretation of tactile signals and the performance of concurrent tasks if both involved the same processing code. To ensure the best overall performance in a task set that requires a high degree of both spatial and symbolic processing, the tactification displays were designed to redundantly encode each type of information with both spatial and symbolic patterns. This way, those receiving the tactification signal could employ whichever processing code faced less competition from concurrent task requirements, thus supporting reduced interference at the cognitive processing stage.

The second property of preattentive reference that is being accommodated by the tactile displays that were developed as part of this research effort is to include partial information on potential interruptions to support decision making about whether, and when, to shift attention between tasks. The survey in Chapter 3 helped identify which types of information were most critical for deciding when to switch attention away from ongoing tasks to manage a patient's blood pressure. Chapter 4 then investigated physiological monitoring and multitasking performance when iconic vibrotactile displays communicated the two most important types of information: the level and rate of change in Mean Arterial Pressure. The results of this study suggested that both types of information support physiological monitoring, but that more natural ways to relate the information (than iconic encoding) were needed and subsequently developed and tested during the final design and simulation study.

Finally, the tactification displays were designed to support the third property of preattentive reference – supporting assessment in a mentally economical way – by exploiting natural mappings between the signal modulations and the underlying data. Chapter 5 described in detail how the two tactifications (as well as the tactile alarm) were designed to support natural mappings in different ways. Ultimately, it was found that signals modulations which effectively mapped the salience of the signal to the severity associated with the parameter levels led to improved performance at detecting trends and abnormal levels in those parameters.

The "hybrid" tactification display, which combined the properties of continuously informing displays with those of a multistage graded alarm system, led to the greatest improvements in monitoring and multitask performance, when compared to performance with a display configuration that is common in most current clinical settings. It supported a higher likelihood of detecting serious health events, as well as significantly earlier detection and correction of these events. Each parameter that was communicated via this display was managed more effectively, without negatively impacting performance on ongoing tasks. Finally, the hybrid display outperformed the alarm and continuous displays in objective measures of multitask performance. This display can serve as a model for continuously-informing displays to support the attention and task management of anesthesiologists, as well as operators in other complex, data-rich domains, such as process control, aviation, or surface transportation.

While objective performance measures generally improved with the continuously-informing tactile displays, it should be noted that this type of display was not universally well-received. Many participants preferred the tactile alarm display, which was much simpler and presented vibration signals less frequently. Surprisingly, a few other participants favored the baseline display configuration, preferring no vibrotactile presentations at all. This illustrates an issue with introducing more data displays to an environment: one cannot assume that "more" is equivalent to "better". While a redistribution of data among the sensory channels may improve parallel processing abilities, the addition of more data sources – even the redundant display of data in a separate modality – may lead to a greater risk of data overload. This risk might be reduced via mechanisms that control the amount of data displayed, and the media in which they are displayed, to ensure the most efficient communication.

In addition to informing display and interface design, this body of work also makes a significant contribution to our understanding of tactile and multimodal information processing. In particular, the work highlights perceptual and cognitive limitations that have not been or are not commonly considered in tactile and multimodal display design. The study described in Chapter 2: Part A introduced and demonstrated a new form of tactile change blindness. It showed how performance in detecting changes in vibration intensity was significantly better with discrete step changes, as compared to gradual, analog changes. Importantly, it also showed how the most complex "transient" event – a mudsplash presentation, led to considerably worse performance when a secondary task load as imposed. This finding was then confirmed in the tactification evaluation study, which was conducted under a considerable secondary task load. In this study, the "continuous" display, which required detecting changes in intensity and spatial location to recognize developing health events, did not support this recognition as well as the hybrid display. This suggests that the "beat pulse" – a modulation of both intensity and temporal properties – may be more resilient to the change blindness effects.

Multimodal information processing and interface design can also be informed by the study described in Chapter 2: Part B, which demonstrated how the benefits of presenting task-relevant stimuli to separate presentation modalities depend on an interaction with processing code. The study showed how two basic ways to encode information into a tactile signal – via spatial or temporal patterns – can result in very different multitasking performance levels that depend heavily on whether primarily spatial or nonspatial/symbolic processing is required by concurrent visual tasks. By quantifying the performance effects of processing code interference between tasks, the findings of this study suggest the extent to which concurrent task processing demands should be considered in tactile interface design.

Finally, the findings from this research also help inform future developments of models of the structure of attentional and processing resources which can be used to describe multitasking performance, such as Multiple Resource Theory (Wickens, 2002; 2008) and Queueing Network-Model Human Processor (e.g., Liu, Feyen, & Tsimoni, 2006). Of particular interest for these models may be the differences in performance between participants, found in both of the studies described in Chapters 4 and 5. Some participants showed substantially improved performance with the tactification displays while others expressed a feeling of "sensory overload". In particular, these findings speak to the issue of a "red-line" of cognitive workload (Grier et al., 2008), which might depend considerably on the number of modalities concurrently engaged in processing task-relevant data. Also, this work can contribute to expansions of predictive models of the allocation of attentional resources, such as SEEV and its extensions (Salience, Effort, Expectancy, Value; e.g., Wickens, Goh, Horrey, Helleberg, & Talleur, 2003; Wickens, McCarley, Steelman-Allen, Sebok, Bzostek, & Sarter, 2009). An expanded multimodal version of this model, for example, may consider how the salience differences, informativeness, and effort involved in processing the different tactification signals, iconic tactile signals, and tactile alarms affected the likelihood of the signals being adequately attended to and processed.

Additionally, the quantification of performance measures, such as the multitask performance metrics employed in Chapter 2: Part B (Equation 2.1) and Chapter 5 (Equation 5.1) may provide valuable empirical input to computational versions of these models, which rely on quantified interference coefficients to calculate predicted overall performance. Valid estimations of such coefficients for particular task sets and contexts can improve the fidelity and predictive power of these models. The models can then be used as tools for informing interface design, for example, to identify encoding methods that can be expected to minimize perceptual and processing code interference.

As in every research effort, this dissertation highlights outstanding questions and suggests directions for future research. These represent some of my future research plans, which I will begin in January 2011 as an assistant professor of Industrial & Systems Engineering at Texas A&M University.

First, the decision of which encoding methods to employ in the design of tactile displays often involves a tradeoff between two important characteristics: the degree of interference a method imposes on cognitive processing resources, and how well the method supports natural mapping of the signal to the represented data. A quantification of the former characteristic was developed in the studies described in Chapter 2: Part B. The development of a similar quantifiable metric for "naturalness" may allow the development of a more complete quantitative model which can be used to develop a tool to predict performance with various display designs during the design process.

Secondly, when concurrently processing even very simple signals via 3 modalities (vision, audition, and touch) some participants in the studies described here reported a state of "sensory overload" which inhibited their performance. These reports suggest a possible interaction between these three modalities that deserves a more controlled investigation. Determining the root causes of this overload effect will be an important step toward the advancement of tactile interfaces in environments that are likely to impose significant loads on the visual and auditory channels.

Third, further improvements may be made to the tactification designs. The discrepancy between performance levels and subjective rankings of the displays suggested a general dislike of frequent vibration presentations. In an effort to make tactification displays more acceptable in real-world environments, adaptive and adaptable mechanisms will be investigated. By supporting context-sensitive or operator-controlled adjustment of the format or frequency of tactile presentations, performance may be best supported and the likelihood of their using these displays may increase.

Fourth, two limitations of the evaluation study described in Chapter 5 suggest additional questions which will be explored in the evaluation of future tactification designs. The first is the likely presence of a learning effect, evidenced by faster induction task completion times in later scenarios, and possibly also indirectly affecting physiological monitoring task performance. Longer training times, possibly over the course of several days or longer, may be necessary to overcome these effects. Additionally, extended training with the tactifications could help determine whether more familiarity with the tactification displays might improve one's ability to "tune out" the signal, allowing it to more effectively support a state of peripheral awareness. The other limitation was the task priority imposed in the evaluation study. Participants were instructed to treat the physiological monitoring task as equivalently important to the induction task, which largely consisted of steps to secure the patient's airway. This instruction conflicted with the task priorities anesthesiologists are trained to follow in clinical practice (for which securing the airway and assuring normal breathing are the stated top priorities). An experimental design that reflects a more natural task priority may generate results that are more predictive of performance in the real clinical setting.

Finally, tactile interface design represents just one direction for supporting improved attention and task management in the clinical setting. Other non-(focal)visual channels, such as audition, peripheral vision, and even olfaction offer unique affordances which can be exploited to effectively guide visual attention and/or support awareness of physiological data without (necessarily) engaging the focal visual channel. The adaptive/adaptable mechanisms mentioned previously in the context of tactification design might also be applied to a multimodal interface. For example, these mechanisms could support the attentional transition between states of "peripheral awareness" and focused attention by transitioning a signal between display modalities.

The efforts described here, and future efforts in the development of tactification and multimodal displays, show promise for reducing the likelihood of monitoring errors and other preventable medical errors. Through continued efforts such as these, we can realize a reduction in healthcare costs associated with preventable medical errors, and, most importantly, improved patient health outcomes.

REFERENCES

Cooper, J.B., Newbower, R.S., Long, C.D., & McPeek, B. (1978). Preventable anesthesia mishaps: A study of human factors. *Anesthesiology, 49*. 399-406.

Fitch, W.T. & Kramer, G. (1994). Sonifying the body electric: Superiority of an auditory over a visual display in a complex, multivariate system. In Kramer, G. (Ed.), *Auditory Display: Sonification, Audification, and Auditory Interfaces*. New York: Addison-Wesley.

Grier, R., Wickens, C., Kaber, D., Strayer, D., Boehm-Davis, D., Trafton, J.G., & St. John, M. (2008). The red-line of workload: Theory, research, and design. *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting*, 1204-1208.

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

Imhoff, M. and Kuhls, S. (2006). Alarm algorithms in critical care monitoring. *Anesthesia and Analgesia*, 102, 1525-1537.

Kohn, L.T., Corrigan, J.M. and Donaldson M.S. (2000). *To err is human: Building a safer health system.* National Academy Press, Washington, DC.

Kramer, G. (1994). Some organizing principles for representing data with sound. In Kramer, G. (Ed.), *Auditory Display: Sonification, Audification, and Auditory Interfaces*. New York: Addison-Wesley. 79-94.

Liu, Y., Feyen, R., & Tsimoni, O. (2006). Queueing network-model human processor (QN-MHP): A computational architecture for multitask performance in human-machine systems. *ACM Transactions on Computer-Human Interaction (TOCHI), 13*(1), 37-70.

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

Sanderson, P.M., Watson, M.O., Russell, W.J., Jenkins, S., Liu, D., Green, N., Llewelyn, K. , Cole, P., Shek, V., & Krupenia, S.S. (2008). Advanced auditory displays and head-mounted displays: Advantages and disadvantages for monitoring by the distracted anesthesiologist. *Anesthesia Analgesia, 106*, 1787-1797.

Seagull, F.J. and Sanderson, P.M. (2001). Anesthesia alarms in context: An observational study. *Human Factors*, 43 (1), 66-78.

Seagull, F.J., Wickens, C., & Loeb, R.G. (2001). When is less more? Attention and workload in auditory, visual, and redundant patient-monitoring conditions. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting*, Santa Monica, CA, 1395-1399.

Trafton, J.G., & Monk, C.A. (2008). Task interruptions. In Boehm-Davis, D.A. (Ed.), *Reviews of human factors and ergonomics, Vol. 3*. Santa Monica, CA: Human Factors and Ergonomics Society. 111-126.

Ulrich, R.S., Zimring, C., Zhu, X., Dubose, J., Seo, G-B., Choi, Y-S., Quan, X., & Joseph, A. (2008). A review of the research literature on evidence-based healthcare design (Part 1). *Health Environments Research & Design, 1*(3), 61-125.

Walsh, T. & Beatty, P.C.W. (2002). Human factors error and patient monitoring. *Physiological Measurement, 23*, R111-R132.

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

Webb, R.K., van der Walt, J.H., Runciman, W.B., Williamson, J.A., Cockings, J., Russell, W.J., and Helps, S. (1993). Which monitor? An analysis of 2000 incident reports. *Anaesthesia and Intensive Care, 21*, 529-542.

Wickens, C.D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Sciences, 3*, 159-177.

Wickens, C.D. (2008). Multiple resources and mental workload. *Human Factors, 50*(3), 449-455.

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.

Wickens, C., McCarley, J., Steelman-Allen, K., Sebok, A., Bzostek, J., & Sarter, N. (2009). NT-SEEV: A model of attention capture and noticing on the flight deck. *Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting*, 769-773.

Woods, D.D. (1995). The alarm problem and directed attention in dynamic fault management. *Ergonomics*, 38 (11), 2371-2393.

Appendix 1: Complete List of Online Survey Questions

1** (mandatory questions labeled with red asterisks)

Please approximate your experience as a practicing Anesthesia Provider. Choose from the drop-down list below.

Please select one...

- Please select one...
- Not an anesthesia provider (please explain below)
- Resident CA-1 (currently)
- Resident CA-2
- Resident CA-3
- CRNA
- Fellow
- Faculty: less than 10 years
- Faculty: 10 years or more

If not a practicing anesthesia provider, please briefly explain your experience with the field.

1a**

You have indicated that you are a resident. Have you completed a vascular rotation?

- Yes
- No
- Currently in vascular rotation - please explain:

Blood pressure is one promising candidate for a continuous tactile display because of its highly dynamic nature, the criticality of maintaining blood pressure within a safe range, and problems associated with current threshold-based auditory alarms (such as the high number of false alarms, requirement to set threshold levels which are ill-defined, and no alarm for critical BP changes if they do not exceed a preset threshold).

2**

In the text box below, please describe at least one blood pressure dynamic/pattern which is especially difficult to monitor, or requires more of your attention.

3**

Please briefly describe a *patient co-morbidity* where management of blood pressure is especially challenging, and explain why.

4**

Do any of the following routine tasks interfere with your ability to closely monitor critical patient parameters such as blood pressure?

Please rate each task listed below. You will have a chance to elaborate later.

Task	**1 No interference**	**2 Slight interference**	**3 High interference**
Setting up infusion pumps or replacing saline drip	☐	☐	☐
Documentation/entering records on Centricity (such as drugs administered)	☐	☐	☐
Administering a drug via syringe	☐	☐	☐
Draining Foley catheter bag	☐	☐	☐
Checking/replacing sensors on patient (such as pulse ox)	☐	☐	☐
Drawing blood from vein or A-line	☐	☐	☐
Intubation	☐	☐	☐
Train-of-4	☐	☐	☐
Preparing drug syringes for the current or a later case	☐	☐	☐

---

5 (optional)

If you can think of any other tasks that interfere with monitoring, please list them below.

Over the next few pages, you will choose **three** tasks which present the most interference with patient monitoring.

6**

Task 1: this task ***most*** interferes with your ability to monitor patient status. (please select one from the dropdown list)

7

Please briefly elaborate. Why does performing this task pose a challenge for patient monitoring?

8

Are there any techniques you use to work around the challenges you listed in the previous question?

9**

Task 2: this is the ***second-most*** interfering task. (please select from the dropdown list)

10

Please briefly elaborate. Why does performing this task pose a challenge for patient monitoring?

11

Are there any techniques you use to work around the challenges you listed in the previous question?

12**

Task 3: this is the ***third-most*** interfering task. (please select from the dropdown list)

13

Please briefly elaborate. Why does performing this task pose a challenge for patient monitoring?

14

Are there any techniques you use to work around the challenges you listed in the previous question?

15

Are there any other possible distractions that may affect your ability to monitor patient status (for example, discussions with surgeons/OR staff or attending anesthesiologists)? If so, please briefly describe these distractions and any techniques you have developed to cope with them.

Consider the task of maintaining blood pressure within a specified range during a Carotid Endarterectomy (CEA). The target BP range will change throughout the procedure (e.g., when the surgeons clamp/unclamp the Carotid artery).
For this task, you will rate each type of information listed below according to the following options:

1: essential/critical to have this info displayed at all times	**2**: essential/critical to display only at certain times	**3**: display is desired if feasible, but not absolutely necessary	**4**: display of information is not necessary or desired

For your ratings, assume a routine case (no unusual patient circumstances) and a single, instantaneous display of information (i.e., no history information is available for any of these parameters).

Feel free to elaborate after any response (especially those with a 2 or 3 rating) in the available comment box.

16**

Please rate according to the scale explained above and elaborate if appropriate.

a. Systolic BP

b. Diastolic BP

c. Mean BP

d. Direction of change of BP (up or down)

e. Rate of change of BP

f. When BP is in a steady state (no change)

g. BP dramatically drops or skyrockets

h. How close the current BP is to the preset alarm threshold limits

i. When BP is outside of preset high/low BP threshold limits

j. How far BP is outside of preset high/low BP threshold limits

---

17

In the boxes below, please list any other types of information which were not included above but which might be beneficial. You will have a chance to rate them on the next page.

k. write-in 1

l. write-in 2

m. write-in 3

17 (continued)**

Please rate the importance of each write-in information type and elaborate if appropriate.

18**

Given a **routine case**, what would be the three most useful types of BP information to display? Please choose from each drop-down list below, in order of usefulness.

Most useful — Please select...

Second-most useful

Third-most useful

Please select...
Systolic BP
Diastolic BP
Mean BP
Direction of change of BP (up or down)
Rate of change of BP
When BP is in a steady state (no change)
BP dramatically drops or skyrockets
How close the current BP is to the preset alarm threshold limits
When BP is outside of preset high/low BP threshold limits
How far BP is outside of preset high/low BP threshold limits
write-in 1: $writein1
write-in 2: $writein2
write-in 3: $writein3

19

Can you think of critical incidents where the set of most useful blood pressure information types may be different from the most useful information for a routine case?

If so, please describe the critical incidents in the boxes on the left, and edit your choices for most useful information types.

Critical Incidents	Most useful information		
CI #1:	(1):	(2):	(3):
CI #2:	(1):	(2):	(3):
CI #3:	(1):	(2):	(3):

19b

Please note if any of the types of BP information you listed as useful in the previous questions are not currently displayed in the OR, and elaborate if appropriate.

The next few pages are going to ask about how to present the three information types which you determined would be most useful for routine cases:
Infotype 1: **$infotype1**
Infotype 2: **$infotype2**
Infotype 3: **$infotype3**

This information will be presented via a ***vibration display***, which will be an arrangement of one or multiple small, vibrating devices affixed to a harness which is worn on the body. The information will be encoded in complex patterns of vibrations.

20a**

For the info type: **$infotype1**, should the frequency of presentation be preset or adjustable by anesthetic personnel?

- ☐ Preset
- ☒ Adjustable

Please elaborate, if appropriate

20b**

Considering the information will be displayed *via vibration patterns*, should it be periodically presented throughout the case, or only during a potential BP problem?

- ☐ throughout the case
- ☐ only during potential BP problem
- ☐ it depends...(please explain):

20c**

During such a problem, should this info be presented (as vibrations) for as long as the problem persists, or only at the start of the potential problem?

- ☐ as long as problem persists
- ☐ only at start of problem
- ☐ it depends...(please explain):

20d**

If the display will not present this information for the entire time that the problem persists, would you prefer to have a vibratory notification once the problem is resolved?

- ☐ yes
- ◉ no notification needed
- ☐ it depends...(please explain):

21a**

For the info type: **$infotype2**, should the frequency of presentation be preset or adjustable by anesthetic personnel?

- ☐ Preset
- ◉ Adjustable

Please elaborate, if appropriate

21b**

Considering the information will be displayed *via vibration patterns*, should it be periodically presented throughout the case, or only during a potential BP problem?

- ☐ throughout the case
- ☐ only during potential BP problem
- ☐ it depends...(please explain):

21c**

During such a problem, should this info be presented (as vibrations) for as long as the problem persists, or only at the start of the potential problem?

- [ ] as long as problem persists
- [ ] only at start of problem
- [ ] it depends...(please explain):

21d**

If the display will not present this information for the entire time that the problem persists, would you prefer to have a vibratory notification once the problem is resolved?

- [ ] yes
- [x] no notification needed
- [ ] it depends...(please explain):

22a**

For the info type: **$infotype3**, should the frequency of presentation be preset or adjustable by anesthetic personnel?

- [ ] Preset
- [x] Adjustable

Please elaborate, if appropriate

22b**

Considering the information will be displayed *via vibration patterns*, should it be periodically presented throughout the case, or only during a potential BP problem?

- [ ] throughout the case
- [ ] only during potential BP problem
- [ ] it depends...(please explain):

22c**

During such a problem, should this info be presented (as vibrations) for as long as the problem persists, or only at the start of the potential problem?

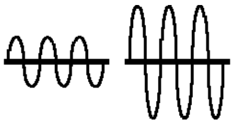
- ( ) as long as problem persists
- ( ) only at start of problem
- ( ) it depends...(please explain):

22d**

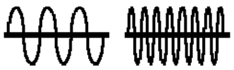
If the display will not present this information for the entire time that the problem persists, would you prefer to have a vibratory notification once the problem is resolved?

- ( ) yes
- (•) no notification needed
- ( ) it depends...(please explain):

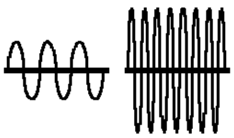
Below are five simple modulations which may be made to a vibration signal in order to encode information.



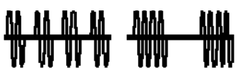
1. Amplitude: displacement of the vibrating device can be more/less



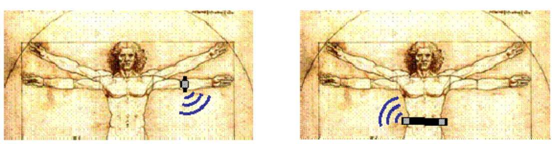
2. Frequency: vibrations can be slower or faster



3. Intensity: simultaneous changes in both the frequency and amplitude of the vibration



4. Pulse frequency/rhythm: vibration 'buzzes' can be longer or shorter and have more or less time between them. Also the buzzes could follow a characteristic pattern or rhythm



5. Body location: with multiple vibrating devices, single vibrations presented to different parts of the body or vibration sequences can be used to represent the nature of certain info types

---

23**

A vibration signal is most effective if the information it represents can be interpreted easily and intuitively. For each of the three info types below, please choose which modulation would be most effective, in your opinion. If there does not appear to be an appropriate option, select 'Other' and suggest your own, if possible.

a. **$infotype1**

- amplitude
- frequency
- intensity
- pulse frequency/rhythm
- body location
- other, please specify:

b. **$infotype2**

- amplitude
- frequency
- intensity
- pulse frequency/rhythm
- body location
- other, please specify:

c. **$infotype3**

- amplitude
- frequency
- intensity
- pulse frequency/rhythm
- body location
- other, please specify:

The vibration display will consist of a small number of vibrating devices affixed to a comfortably-fitting harness. Please rate the following body locations for vibration presentation according to the terms below.

24**

Obstructiveness: the amount of physical interference with other tasks which would result from wearing a harness and being presented with vibrations at this body site.

Body part	**1**	**2**	**3**	**4**	**5**	**6**	**7**
	Least					Most	
Wrist	☐	☐	☐	☐	☐	☐	☐
Forearm	☐	☐	☐	☐	☐	☐	☐
Upper Arm	☐	☐	☐	☐	☐	☐	☐
Thigh	☐	☐	☐	☐	☐	☐	☐
Waist (Front)	☐	☐	☐	☐	☐	☐	☐
Lower Back	☐	☐	☐	☐	☐	☐	☐

25**

Annoyance: the amount of frustration and/or physical discomfort which could be expected from periodic vibrations at the body site.

Body part	**1**	**2**	**3**	**4**	**5**	**6**	**7**
	Least					Most	
Wrist	☐	☐	☐	☐	☐	☐	☐
Forearm	☐	☐	☐	☐	☐	☐	☐
Upper Arm	☐	☐	☐	☐	☐	☐	☐
Thigh	☐	☐	☐	☐	☐	☐	☐
Waist (Front)	☐	☐	☐	☐	☐	☐	☐
Lower Back	☐	☐	☐	☐	☐	☐	☐

26**

Distraction: how likely your train of thought would be interrupted during ongoing tasks when periodic vibrations are presented to the body site.

Body part	**1**	**2**	**3**	**4**	**5**	**6**	**7**
	Least					Most	
Wrist	☐	☐	☐	☐	☐	☐	☐
Forearm	☐	☐	☐	☐	☐	☐	☐
Upper Arm	☐	☐	☐	☐	☐	☐	☐
Thigh	☐	☐	☐	☐	☐	☐	☐
Waist (Front)	☐	☐	☐	☐	☐	☐	☐
Lower Back	☐	☐	☐	☐	☐	☐	☐

Thank you very much for your input in this survey. The following questions are optional, meant to give you a chance to give any other ideas you might have for vibration displays in the OR. Your input here will contribute to the next round of study with this display type.

27a

Are there physiological parameters besides blood pressure that may be equally good candidates for the use of a tactile vibration display?

27b

A tactile display could also be used as a reminder of tasks/actions that need to be completed during the current case or for upcoming cases. Please identify some actions for which a reminder would be highly desirable and useful.

27c

Are there any other uses for a vibrotactile device, inside or outside the OR, which you would consider helpful?

27d

We would appreciate any other comments, suggestions, or ideas you may have related to this project.

Appendix 2: TAPIS Patient Chart

Name: Propono, Tactus DOB: 4/17/1947
ID: 123456789-0035

63 yo M

Wt: 85 kg

Ht: 68 in

Blood Type: O neg

Procedure: Carotid Endarterectomy
Severe reflux: RSI recommended

Previous Medical History:

- TIA: Transient Ischemic Attack (mild stroke) 2 weeks ago, no lasting health effects.
- Previous motor vehicle accident 25 years ago. Surgically repaired clavicle fracture, no lasting effects.
- Hypertension, runs at 130/85. Treated with Atenolol 50mg OD which he took this morning.
- Family history of CAD (Coronary Artery Disease), no prior MI and a normal ECG.
- Moderate smoker for 45 years, no prior treatment of respiratory disease.

Allergies: penicillin, dairy

Pt last ate 0800 hours ago and drank water 0300 hours ago

ADVISE Maintain:

- MAP ≈ 100 (+/- 10) mmHg
- EtCO2 ≈ 38 (+/- 4) mmHg
- TV ≈ 500 (+/- 50) mL

Appendix 3: TAPIS Induction Task Steps

Scenario-specific events are highlighted, as follows:

a. SCENARIO1:
- Sensoroff2 = 1
- Hypo_prop = 1
- Hyper_scope2 = 1
- Hyperventilation = 1
- Sensoroff6 = 1
- Pneumothorax = 1
- Obstruction = 1

b. SCENARIO 2:
- Hyper_vomit = 1
- Reflux1 = 1
- Hypovolemia2 = 1
- Sensoroff4 = 1
- RMI1 = 1
- Hypercarbia = 1
- Anaph = 1

c. SCENARIO 3:
- Hyper_nervous = 1
- Shortshallow2 = 1
- Hypovolemia1 = 1
- Sensoroff3 = 1
- Inad_paralysis = 1
- RMI2 = 1
- Sensoroff5 = 1

d. SCENARIO 4:
- Sensoroff1 = 1
- Shortshallow1 = 1
- PEA = 1
- Hyper_scope1 = 1
- Reflux2 = 1
- Hypo_iso = 1
- Hypoventilation = 1

Induction Steps:

1) Patient has arrived
    a. sim starts with the patient in room
    b. start baseline script: 1, 2, 3, or 4, depending on scenario
2) Begin saline drip (open 1/3)
3) Set up the infusion pumps: recommend remifentanil 0.05 mcg/kg/min, phenylephrine 10 mcg/min. Press A to continue to next step.
4) Turn to tool table to set up mask ventilation
5) Pick up mask and attach it to ventilator hose
6) Turn on gas flow of 100% O2, 8 Lpm
    a. If Sensoroff1 = 1: run Sensors script
7) Place mask on face
    a. start Vent_normal script
    b. If Hyper_nervous = 1: Begin Hyper_nervous script
8) Hold mask on face: hold wiimote upright and press "A" button. Preoxygenate for 1 minute
    a. If Shortshallow1 = 1: Begin Vent_shortshallow1 script
    b. 40 seconds in:
        ▪ If Shortshallow2 = 1: Begin Vent_shortshallow2 script
9) Administer lidocaine while maintaining mask seal
    a. If Sensoroff2 = 1: run Sensors script
10) While maintaining mask seal, administer propofol
    a. If Hypo_prop = 1: begin hypo_prop script
11) While maintaining mask seal, administer succinylcholine
12) Maintain mask seal, you have instructed nurse to apply cricoid pressure
    a. If PEA = 1: begin PEA arrest scripts
    b. Wait 5 seconds then advance to next step
13) While maintaining mask seal, wait 30 seconds for drugs to take effect
    a. If Sensoroff3 = 1: run Sensors script
    b. Begin Vent_propofol script
14) Patient is unconscious, verify mask ventilation by squeezing bag
    a. If Hypovolemia1 = 1: begin Hypovolemia script
    b. Execute subroutine Manual ventilation
15) Remove mask and place on table
16) Turn gas flow rate to 0
    a. If Sensoroff4 = 1: run Sensors script
17) Pick up laryngoscope to examine airway
18) Open mouth
19) Insert scope
    a. Begin Laryngoscopy event script

20) Pull back to displace tongue and expose airway
    a. If Hyper_scope1 = 1: start Hyper_scope script
    b. If Reflux1 = 1, execute "vomit in the airway" subroutine
21) Maintain tongue displacement, pick up ETT and insert into trachea
    a. If Reflux2 = 1, execute "vomit in the airway" subroutine
    b. If Hyper_scope2 = 1: start Hyper_scope script
22) Hold tube in place with mouse and remove laryngoscope
23) Grab and hold tube with wiimote: press and hold "A" with wiimote held upright
24) Inflate endotracheal tube cuff
    a. If Hypovolemia2 = 1: begin Hypovolemia script
    b. If Sensoroff5 = 1: run Sensors script
    c. If Reflux1 = 1 or Reflux2 = 1: "Pass suction into ETT to remove any gastric contents from airway"
        - Execute pass suction subroutine
25) Hook up tube to ventilator hose
26) Squeeze bag for 3 breaths and verify placement with EtCO2
    a. If Inadequate_depth = 1:
        - Begin Vent_inadequate ventilation script and Inadequate_depth script
        - Patient begins to "buck"
    b. If Sensoroff6 = 1: run Sensors script
    c. Execute Manual ventilation subroutine
27) Turn up flow rate, turn on Isoflurane, 1.0 %, and turn on mechanical ventilator
    a. When ventilator turned on, start MechVent_baseline script: 1, 2, 3, or 4, depending on scenario
    b. If Inadequate_depth = 1
        - Stop Vent_inadequate ventilation script
        - Stop Inadequate_depth script
        - Start Inad_depth_Mech
        - Start MechVent_inadeaquate script
    c. If RMI1 = 1, run MechVent_RMI script until RMI = 0
    d. If Hyperventilation = 1: start MechVent_hyperventilation event script
    e. If Hypo_iso = 1: begin Hypo_iso script
28) Auscultate lungs
    a. Execute Auscultate subroutine
29) Tape tube in place
30) Verify paralysis with train-of-4
    a. If Hypercarbia = 1: begin MechVent_hypercarbia script
    b. Execute Train-of-4 subroutine
    c. If Pneumothorax = 1: begin pneumothorax script and MechVent_pneumothorax

31) Give Vecuronium
    a. If Anaph = 1: begin running anaphylaxis script and MechVent_anaphylaxis
32) Pick up esophageal temperature probe from table and insert into esophagus
    a. If RMI2 = 1: run MechVent_RMI event script until RMI2 = 0
    b. If Hypoventilation = 1: start MechVent_hypoventilation script
33) Tape probe in place and tape eyes closed
34) You have completed anesthesia induction and the surgeon is now beginning surgery. Document each induction step on the Centricity display, recording the approximate time each step was completed.
    a. About 30 seconds in:
    b. If Anaph = 1:
        - patient develops truncal rash over 1 minute (visually turns pinkish around upper chest)
    c. If Obstruction = 1, send signal to Clif's to start MechVent_obstruction script, interrupt when obstruction = 0

SUBROUTINES: these can be activated from the TROUBLESHOOTING MENU or via the normal script

35) Draw blood subroutine task
36) Auscultate subroutine
37) Train of 4 subroutine
38) Manual ventilation subroutine
39) Adjust ETT subroutine
40) Pass suction subroutine
41) Vomit in the airway subroutine
    a. If Hyper_vomit = 1: Begin Hyper_vomit script
42) Decompression subroutine
43) CPR subroutine
44) Administer drug subroutine
45) Call for help subroutine
46) Reassure patient subroutine

Appendix 4: Normal Fluctuation Script Example

Time	ECG	SYS	DIA	SPO2	TMP1	Ppeak	EtCO2	TV
	75	130	85	99	37.6	15	49	500
0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0
2	0	0	0	1	0.1	0	0	0
3	0	1	1	0	0	0	0	0
4	1	0	0	-1	0	0	-2	-10
5	1	0	0	0	0.1	0	0	0
6	0	1	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	-3	-11
9	0	0	0	1	0	0	0	0
10	1	-1	-1	0	0.1	0	0	0
11	0	0	0	0	0	0	0	0
12	-1	0	0	0	-0.1	0	-2	24
13	0	0	0	0	0	0	0	0
14	0	1	1	0	0	0	0	0
15	1	0	-1	0	0	0	0	0
16	0	1	1	0	0	0	-3	13
17	0	0	0	0	0	0	0	0
18	-1	0	0	0	0	0	0	0
19	-1	-1	-1	0	0	0	0	0
20	0	0	1	-1	0	0	-1	7
21	0	0	0	0	0	0	0	0
22	1	-1	-1	0	0	0	0	0
23	0	0	0	0	0	0	0	0
24	0	0	1	0	-0.1	0	1	-22
25	0	0	0	0	0	0	-1	0
26	1	0	0	0	0	0	0	0
27	0	1	0	0	0	0	0	0
28	1	0	0	0	0	0	-1	-13
29	0	0	0	0	0	0	0	0
30	0	0	1	0	0	0	0	0
31	0	-1	-1	0	0	0	-1	0
32	0	0	1	0	0	0	1	16
33	0	0	0	0	0	0	0	0
34	-1	0	0	0	0	0	0	0
35	-1	-1	-1	0	0	0	0	0
36	-1	0	0	0	0	0	0	-24
37	0	0	0	-1	-0.1	0	0	0
38	0	1	0	0	0	0	0	0
39	0	0	0	0	0	0	1	0

40	0	1	0	0	0.1	0	0	-3
41	0	0	0	0	0	0	0	0
42	0	0	0	0	0	0	0	0
43	0	-1	-1	0	0.1	0	0	0
44	1	0	0	0	0	0	0	-2
45	0	0	0	0	0	0	0	0
46	0	0	0	0	0	0	0	0
47	0	-1	-1	1	0	0	0	0
48	1	-1	-1	0	0.1	0	0	17
49	0	-2	-1	0	0	0	0	0
50	0	-1	0	0	-0.1	0	0	0
51	0	-1	-1	0	0	0	-1	0
52	1	0	0	0	0	0	0	-9
53	1	0	0	0	0	0	0	0
54	0	0	0	0	0	0	1	0
55	0	0	0	0	0	0	0	0
56	-1	0	0	0	0	0	0	27
57	0	0	0	0	0	0	-1	0
58	0	0	0	0	0	0	0	0
59	0	1	1	-1	0	0	0	0
60	-1	0	0	0	0	0	0	14
61	-2	1	1	0	0	0	0	0
62	0	0	-1	0	-0.1	0	0	0
63	-1	0	0	0	0	0	0	0
64	0	0	0	0	0	0	0	-24
65	0	-1	-1	0	0	0	1	0
66	0	0	0	0	0	0	0	0
67	0	0	0	0	0	0	0	0
68	-1	1	1	0	0	0	0	6
69	0	0	1	0	0	0	0	0
70	0	1	1	0	0	0	1	0
71	0	0	0	0	0	0	-1	0
72	0	0	0	0	0	0	0	9
73	-1	0	0	0	0	0	0	0
74	0	0	0	0	0	0	0	0
75	0	0	0	0	-0.1	0	0	0
76	0	0	0	0	0	0	0	-14
77	0	0	0	0	0	0	1	0
78	0	1	0	0	0	0	0	0
79	0	0	0	0	0	0	0	0
80	1	-1	-1	0	0	0	0	-4
81	0	1	0	1	0	0	0	0
82	0	0	0	0	0	0	0	0
83	0	0	1	0	0	0	0	0
84	0	0	0	0	0	0	0	15
85	1	0	0	0	0	0	0	0
86	1	1	0	0	0	0	0	0

87	1	-1	-1	0	0	0	0	0
88	0	0	0	0	0	0	0	-27
89	0	0	0	0	0	0	0	0
90	0	0	-1	0	0	0	0	0
91	0	-1	-1	0	-0.1	0	0	0
92	-1	-2	0	0	0	0	1	-5
93	-2	0	0	0	0	0	0	0
94	0	1	1	0	0	0	0	0
95	-1	0	0	-1	-0.1	0	0	0
96	-2	0	0	0	0	0	0	31
97	0	0	0	0	0	0	-1	0
98	0	1	1	0	0	0	0	0
99	1	1	1	0	0	0	0	0
100	0	1	0	0	-0.1	0	0	11
101	0	0	0	0	0	0	0	0
102	0	0	0	0	0	0	0	0
103	0	1	0	0	0.1	0	0	0
104	1	0	0	0	0	0	0	-11
105	0	0	1	0	0	0	0	0
106	0	0	0	0	0	0	0	0
107	1	1	0	0	0	0	0	0
108	0	0	0	0	0	0	1	-15
109	0	2	1	0	0	0	0	0
110	0	-1	2	-1	0	0	0	0
111	2	1	1	0	0	0	0	0
112	1	0	-1	0	0	0	0	7
113	1	0	0	0	0	0	0	0
114	0	0	0	0	0	0	-1	0
115	0	1	0	0	0.1	0	0	0
116	0	-1	0	0	0	0	0	9
117	0	0	0	0	0	0	0	0
118	-1	0	0	0	0	0	0	0
119	0	0	0	0	0	0	0	0
120	0	-1	1	0	0	0	0	-4

(only first 120 s shown of a 1500 s script)

Appendix 5: Example Scripts for Blood Pressure, Respiration, and Emergency Events

Blood Pressure Event Example: Hyper_nervous

Time	ECG	SYS	DIA	SPO2	TMP1
0	0	0	0	0	0
1	1	1	0	0	0
2	2	2	1	0	0
3	1	3	3	0	0
4	1	3	2	0	0
5	2	1	3	0	0
6	0	0	1	0	0
7	1	0	0	0	0
8	0	0	0	0	0
9	1	0	0	0	0
10	0	0	-1	0	0
11	1	-1	-1	0	0
12	1	0	2	0	0
13	-1	0	0	0	0
14	0	0	0	0	0
15	0	1	1	0	0
16	0	2	2	0	0
17	0	2	1	0	0
18	0	1	1	0	0
19	0	2	1	0	0
20	0	1	1	0	0
21	0	1	2	0	0
22	0	2	3	0	0
23	0	3	1	0	0
24	0	0	1	0	0
25	0	1	0	0	0

Respiration Event Script Example: Hyperventilation

Time	Ppeak	EtCO2	TV
0	0	0	0
1	0	0	0
2	0	0	0
3	0	0	0
4	0	-2	75
5	0	0	0
6	0	0	0
7	0	0	0
8	0	-2	0
9	0	0	0
10	0	0	0
11	0	0	0
12	0	0	0
13	0	0	0
14	0	0	0
15	0	0	0
16	0	-2	77
17	0	0	0
18	0	0	0
19	0	0	0
20	0	-2	28
21	0	0	0
22	0	0	0
23	0	0	0
24	0	0	37
25	0	0	0

Emergency Event Script Example: Anaphylaxis

Time	ECG	SYS	DIA	SPO2	TMP1	Ppeak	EtCO2	TV
0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0
8	0	-1	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0
10	0	0	-1	0	0	0	0	0
11	0	-1	0	0	0	0	0	0
12	0	0	-1	0	0	0	0	0
13	0	-1	-2	1	0	0	0	0
14	0	-1	0	0	0	0	0	0
15	0	-1	-1	0	0	0	0	0
16	1	-1	-2	0	0	0	0	0
17	1	-1	-1	0	0	0	0	0
18	1	-1	0	-1	0	0	0	0
19	1	0	-1	0	0	0	0	0
20	1	-1	0	0	0	0	0	-19
21	0	0	0	0	0	0	0	0
22	1	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0
24	1	0	0	0	0	0	0	-17
25	1	-1	0	0	0	0	0	0
26	0	0	-1	0	0	0	0	0
27	0	0	0	0	0	0	0	0
28	1	0	0	0	0	0	0	-22
29	0	0	0	-1	0	0	0	0
30	1	0	0	0	0	0	0	0
31	1	0	0	1	0.1	0	0	0
32	0	1	0	0	0	0	0	-17
33	1	-1	1	0	0	0	0	0
34	1	1	-1	0	0	0	0	0
35	-1	-1	0	0	0	0	0	0
36	1	0	-1	0	0	0	1	0

37	0	-1	0	0	0	0	0	0
38	1	0	-1	0	0	0	0	0
39	1	-1	-1	0	0	0	0	0
40	1	-1	-1	0	0.1	0	1	0
41	-1	-2	-1	1	0	0	0	0
42	1	-1	-1	0	0	0	0	0
43	0	-1	-1	0	0	0	0	0
44	0	-2	-2	-1	0	0	0	0
45	1	-1	-1	0	0	0	0	0
46	1	-1	0	0	0.1	0	0	0
47	0	-2	-1	0	0	0	0	0
48	1	-1	-2	0	0	0	1	0
49	0	0	-1	0	0	0	0	0
50	1	-1	-1	-1	0.1	0	0	0
51	1	-1	-1	0	0	0	0	0
52	-1	0	0	0	0	0	1	-17
53	0	-1	0	-1	0	0	0	0
54	1	-1	-1	0	0	0	0	0
55	1	0	0	0	0	0	0	0
56	1	0	0	0	0.1	0	0	-22
57	0	-1	0	-1	0	0	0	0
58	1	-1	0	0	0	0	0	0
59	0	-1	-1	0	0	0	0	0
60	0	0	0	0	0.1	0	0	-23
61	1	-1	-1	0	0	0	0	0
62	1	0	0	0	0	0	0	0
63	0	0	0	0	0	0	0	0
64	1	-1	0	0	0	0	0	-27
65	-1	0	-1	0	0.1	0	0	0
66	1	-1	0	0	0	0	0	0
67	1	0	0	0	0	0	0	0
68	1	-1	0	0	0	0	0	-33
69	0	0	-1	1	0.1	0	0	0
70	0	-1	0	0	0	0	0	0
71	1	-1	0	-1	0	0	0	0
72	1	0	0	0	0	0	-1	-37
73	0	0	0	0	0	0	0	0
74	0	-1	-1	0	0	0	0	0
75	0	-1	0	-1	0.1	0	0	0
76	1	0	-1	0	0	0	-1	-47
77	1	-1	0	0	0	0	0	0
78	1	0	0	0	0.1	0	0	0

79	0	0	0	0	0	0	0	0
80	1	-1	-1	0	0	0	-2	-43
81	0	0	0	0	0	0	0	0
82	1	-1	0	0	0	0	0	0
83	0	0	0	0	0.1	0	0	0
84	1	0	-1	0	0	0	0	-33
85	1	-1	0	0	0	0	0	0
86	0	0	1	0	0	0	0	0
87	0	-1	0	-1	0	0	0	0
88	1	0	0	0	0.1	0	-2	-32
89	0	-1	0	0	0	0	0	0
90	1	0	-1	0	0	0	0	0
91	1	-1	0	0	0	0	0	0
92	0	0	0	0	0	0	-3	0
93	2	-1	-1	0	0	0	0	0
94	1	0	0	-1	0	0	0	0
95	1	-1	0	0	0.1	0	0	0
96	2	0	1	0	0	0	-1	0
97	-1	0	0	0	0	0	0	0
98	1	-1	0	-1	0	0	0	0
99	0	-1	-1	0	0	0	0	0
100	1	-1	0	0	0	0	-1	0
101	1	0	0	0	0	0	0	0
102	1	-1	-1	0	0	0	0	0
103	0	0	-1	-1	0	0	0	0
104	1	0	0	0	0	0	-2	0
105	1	-1	0	0	0	0	0	0
106	0	0	1	0	0.1	0	0	0
107	1	-1	0	0	0	0	0	0
108	1	0	0	0	0	0	-2	0
109	2	-1	-1	0	0	0	0	0
110	-1	0	0	0	0	0	0	0
111	0	-1	1	0	0	0	0	0
112	1	-1	0	0	0	0	0	0
113	1	0	1	0	0	0	0	0
114	2	-1	0	0	0	0	0	0
115	-1	0	0	-1	0	0	0	0
116	0	0	-1	0	0	0	0	0
117	1	-1	0	0	0	0	0	0
118	0	0	0	-1	0.1	0	0	0
119	1	0	1	0	0	0	0	0
120	0	-1	1	0	0	0	0	0

121	1	0	0	0	0	0	0	0
122	0	0	0	0	0	0	0	0
123	-1	0	1	0	0	0	0	0
124	0	1	0	0	0	0	0	0
125	0	0	-1	0	0	0	0	0
126	1	0	0	0	0.1	0	0	0
127	0	-1	0	0	0	0	0	0
128	-1	0	0	0	0	0	0	0
129	0	0	0	0	0	0	0	0
130	1	0	0	0	0	0	0	0
131	1	0	-1	0	0	0	0	0
132	0	0	0	0	-0.1	0	0	0
133	0	1	0	0	0	0	0	0
134	2	0	0	0	0	0	0	0
135	-1	0	0	0	0	0	0	0
136	0	-1	1	0	0	0	0	0
137	0	0	0	0	0	0	0	0
138	-1	-1	-1	0	0	0	0	0
139	0	0	-1	0	0	0	0	0
140	0	0	0	0	-0.1	0	0	0
141	-1	0	0	0	0	0	0	0
142	0	0	0	0	0	0	0	0
143	0	-1	0	0	0	0	0	0
144	1	0	-1	0	0	0	0	0
145	0	0	0	0	0	0	0	0
146	-1	0	0	0	0.1	0	0	0
147	-1	0	1	0	0	0	0	0
148	0	1	0	0	0	0	0	0
149	0	0	0	0	0	0	0	0
150	0	0	1	0	0	0	0	0
151	0	0	0	0	0	0	0	0
152	0	0	0	0	0	0	0	0
153	0	0	0	0	0	0	0	0
154	0	0	0	0	0	0	0	0
155	0	0	0	0	0	0	0	0
156	0	0	0	0	0	0	0	0
157	0	0	0	0	0	0	0	0
158	0	0	0	0	0	0	0	0
159	0	0	0	0	0	0	0	0
160	0	0	0	0	0	0	0	0
161	0	0	0	0	0	0	0	0
162	0	0	0	0	0	0	0	0

163	0	0	0	0	0	0	0	0
164	0	0	0	0	0	0	0	0
165	0	0	0	0	0	0	0	0
166	0	0	0	0	0	0	0	0
167	0	0	0	0	0	0	0	0
168	0	0	0	0	0	0	0	0
169	0	0	0	0	0	0	0	0
170	0	0	0	0	0	0	0	0
171	0	0	0	0	0	0	0	0
172	0	0	0	0	0	0	0	0
173	0	0	0	0	0	0	0	0
174	0	0	0	0	0	0	0	0
175	0	0	0	0	0	0	0	0
176	0	0	0	0	0	0	0	0
177	0	0	0	0	0	0	0	0
178	0	0	0	0	0	0	0	0
179	0	0	0	0	0	0	0	0
180	0	0	0	0	0	0	0	0
181	0	0	0	0	0	0	0	0
182	0	0	0	0	0	0	0	0
183	0	0	0	0	0	0	0	0
184	0	0	0	0	0	0	0	0
185	0	0	0	0	0	0	0	0
186	0	0	0	0	0	0	0	0
187	0	0	0	0	0	0	0	0
188	0	0	0	0	0	0	0	0
189	0	0	0	0	0	0	0	0
190	0	0	0	0	0	0	0	0
191	0	0	0	0	0	0	0	0
192	0	0	0	0	0	0	0	0
193	0	0	0	0	0	0	0	0
194	0	0	0	0	0	0	0	0
195	0	0	0	0	0	0	0	0
196	0	0	0	0	0	0	0	0
197	0	0	0	0	0	0	0	0
198	0	0	0	0	0	0	0	0
199	0	0	0	0	0	0	0	0
200	0	0	0	0	0	0	0	0
201	0	0	0	0	0	0	0	0
202	0	0	0	0	0	0	0	0
203	0	0	0	0	0	0	0	0
204	0	0	0	0	0	0	0	0

205	0	0	0	0	0	0	0	0
206	0	0	0	0	0	0	0	0
207	0	0	0	0	0	0	0	0
208	0	0	0	0	0	0	0	0
209	0	0	0	0	0	0	0	0
210	0	0	0	0	0	0	0	0
211	0	0	0	0	0	0	0	0
212	0	0	0	0	0	0	0	0
213	0	0	0	0	0	0	0	0
214	0	0	0	0	0	0	0	0
215	0	0	0	0	0	0	0	0
216	0	0	0	0	0	0	0	0
217	0	0	0	0	0	0	0	0
218	0	0	0	0	0	0	0	0
219	0	0	0	0	0	0	0	0
220	0	0	0	0	0	0	0	0
221	0	0	0	0	0	0	0	0
222	0	0	0	0	0	0	0	0
223	0	0	0	0	0	0	0	0
224	0	0	0	0	0	0	0	0
225	0	0	0	0	0	0	0	0
226	0	0	0	0	0	0	0	0
227	0	0	0	0	0	0	0	0
228	0	0	0	0	0	0	0	0
229	0	0	0	0	0	0	0	0
230	0	0	0	0	0	0	0	0
231	0	0	0	0	0	0	0	0
232	0	0	0	0	0	0	0	0
233	0	0	0	0	0	0	0	0
234	0	0	0	0	0	0	0	0
235	0	0	0	0	0	0	0	0
236	0	0	0	0	0	0	0	0
237	0	0	0	0	0	0	0	0
238	0	0	0	0	0	0	0	0
239	0	0	0	0	0	0	0	0
240	0	0	0	0	0	0	0	0

Appendix 6: Hybrid Display "Pre-test" Presentations and Participant Scores

Pre-test presentations:

presentation #	BP change	TV change	ETCO2 change	RR change	% of changes correctly reported (across all participants)
1	+20	--	--	--	100
2	--	--	+8	--	100
3	--	-150	--	--	100
4	--	+75	--	--	93.75
5	--	--	--	+25	87.5
6	-10	--	--	--	93.75
7	--	+150	-8	--	93.75
8	-20	-150	--	--	93.75
9	+20	--	-4	--	96.875
10	--	--	-8	+25	78.125
11	--	-75	+4	--	84.375
12	-10	--	+4	--	96.875
13	-20	-150	+8	--	100
14	+20	+75	-4	--	91.75
15	+10	-75	+4	--	91.75

Pre-test scores, participants 1 – 8:

presentation #	1	2	3	4	5	6	7	8
1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	0
5	1	0.5	0.5	1	1	1	1	1
6	1	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1	1
8	1	0.5	1	1	1	1	1	1
9	1	0.5	1	1	1	1	1	1
10	0.5	1	0.5	0.5	1	1	1	1
11	1	1	0.5	1	1	1	1	1
12	1	1	1	1	1	1	1	1
13	1	1	1	1	1	1	1	1
14	1	0.67	0.67	1	0.67	1	1	1
15	0.67	0.67	0.67	1	1	1	1	1

Total score	14.2	12.8	12.8	14.5	14.7	15	15	14
% changes correctly reported	92.6	83.4	83.4	96.3	96.3	100	100	96.3

Pre-test scores, participants 9 – 16:

presentation #	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1
5	0	1	1	1	1	1	1	1
6	1	1	0	1	1	1	1	1
7	1	1	0.5	1	0.5	1	1	1
8	1	0.5	1	1	1	1	1	1
9	1	1	1	1	1	1	1	1
10	1	1	0	1	1	0.5	0.5	1
11	1	1	0.5	0.5	1	0.5	1	0.5
12	1	1	1	1	1	1	0.5	1
13	1	1	1	1	1	1	1	1
14	1	1	0.67	1	1	1	1	1
15	1	1	1	1	1	0.67	1	1

Total score	14	14.5	11.7	14.5	14.5	13.7	14	14.5
% changes correctly reported	96.3	96.3	77.8	96.3	96.3	88.9	92.6	96.3

Appendix 7: NASA-TLX Subjective ratings worksheet

Subject #___________ Scenario___________ Display configuration_____________

Please rate the following by making a mark on the analog scales below.

Mental Demand: How mentally demanding (thinking, deciding, calculating, remembering, searching, etc.) was the complete task set (induction tasks and monitoring/managing the patient's health)?

Very Low Very High

Physical Demand: How physically demanding (changing posture, interacting with controls) was the complete task set?

Very Low Very High

Temporal Demand: How hurried or rushed were you in completing the task set?

Very Low Very High

Performance: How successful were you in performing the task set? (NOTE reversed scale)

Perfect Failure

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

Very Low Very High

Frustration: How insecure, discouraged, irritated, stressed, and annoyed were you during this task set?

Very Low Very High

Appendix 8: Postexperiment Questionnaire

1) In the table below, please rank your experience with each of the four display configurations in the experimental task scenarios (1 = best, 4 = worst). Ties are acceptable.

	Baseline (no tactile display)	Tactile alarm	"Continuous" Tactile display	"Hybrid" Tactile display
Supporting multitasking performance				
Minimizing cognitive demand in multitask set				
Minimizing annoyance				
Minimizing unnecessary distraction (e.g., false alarms)				
Maximizing trust (e.g., in system's ability to capture your attention when situation is critical)				
Comfort				
Overall preference				

If you care to elaborate on any of the above rankings, please do so here.

2) Please describe any notable strategies for performing the multitask set: simultaneously completing the patient interaction tasks (administering drugs, intubation, etc.), monitoring/managing the patient's health, and monitoring the NIRS display.

3) Regardless of display configuration, were there any difficulties you found in detecting and/or finding the proper response to any of the scenario events?

4) Are there any other issues with the simulation that you'd care to note which may have affected your performance?

5) Have you ever been frustrated by the existing patient monitoring displays in UM hospitals? Please explain.

6) Assume the vibrotactile display which you ranked as "best" in terms of overall preference (or second-best, if you ranked the baseline condition as best) in question 1) went through a product development cycle which resulted in a comfortable, lightweight, and wireless version of the one evaluated in this study. On a scale of 1 – 10, how *useful* would you expect such a display to be in a clinical setting? Please explain your rating.

7) How often/in what circumstances would you make use of the display described in 6)?

8) Other than practical improvements such as making the system wireless, are there any other improvements you would recommend be made to a vibrotactile patient monitoring display to improve its usability/usefulness?