Heuristic Evaluation of Data Integration and Visualization Software Used for Continuous Monitoring to Support Intensive Care: A Bedside Nurse's Perspective

Ying Ling Lin1,3, Anne-Marie Guerguerian3,4, Peter Laussen3,4 and Patricia Trbovich1,2,4,*

1Institute of Biomaterials and Biomedical Engineering, University of Toronto, Canada
2Centre for Global eHealth Innovation, University Health Network, Canada
3Interdepartmental Division of Critical Care Medicine, Hospital for Sick Children, Canada
4Faculty of Medicine, University of Toronto, Canada
5Neuroscience and Mental Health Research, Hospital for Sick Children, Canada
6Institute of Health Policy, Management and Evaluation, University of Toronto, Canada

*Corresponding author: Patricia Trbovich, Institute of Biomaterials and Biomedical Engineering, Faculty of Engineering, University of Toronto, Canada, Tel: +1-416-340-4800 ext. 7180; Fax: +1-416-340-3595; E-mail: Patricia.Trbovich@uhn.ca

Received date: Aug 01, 2015; Accepted date: Sep 22, 2015; Published date: Sep 30, 2015

Copyright: © Lin YL, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

The Intensive Care Unit (ICU) is a complex and technologically advanced healthcare setting. Technologies enable continuous monitoring through patient signals that are sensed, recorded and displayed at the bedside. Although such technologies have significantly decreased mortality rates in the ICU, the large amounts of data have contributed to clinician information overload. Critical care nurses spend more than half of their time scanning and assimilating information from disparate monitors, at the bedside to assess the patient status. Software that integrates and allows visualization of large data sets on a single screen are now available. In the present study, we evaluated software entitled T3™ (Tracking, Trajectory and Triggering). Such computationally powerful software has great potential to support nurses’ monitoring and decision-making tasks but the usability, efficiency, and effectiveness of the software are key to end-user adoption. As such, we conducted a Heuristic Evaluation, where the study’s evaluators interacted with the software interfaces and were asked to comment on it by describing the usability issues and if they were in compliance with established usability principles, or heuristics, specifically for medical device interfaces.

A total of 50 usability issues associated with 194 heuristic violations were found. Identified issues included difficulty with choosing the time period of the patient data signals, distinguishing between several patient signals and appearance of patient values which were imperceptible to evaluators; both issues could lead nurses to misinterpret the timing and/or the physiological status of the patient (e.g., time of shock and exact value of vitals). Heuristic evaluation, an efficient and inexpensive method, was successfully applied to the T3™ software to identify usability problems that if left unresolved could lead to patient safety issues. These findings may have broad implications for the design of the T3™ and other continuous monitoring systems.

Keywords: Data integration; Visualization; Heuristic evaluation; Multimodal monitoring display; Intensive care; Nursing; Trend; User interface

Introduction

Intensive care units (ICUs) are settings where close monitoring and interventions aimed at achieving homeostasis (i.e., stable vitals within target ranges) are performed on the most fragile patients. The complexity of a pediatric patient's underlying condition is exacerbated by their rapidly evolving developmental physiology [1]. For example, target ranges for a basic vital such as heart rate is highly dependent on age [2]. Long-term monitoring of the critically-ill, pediatric patient is a signature feature of the intensive care unit, and is often associated with the heavy use of monitoring technologies, which collectively, generate large quantities of data[3]. Clinicians specialized in critical care have been known to experience “information overload” [4,5] due to a high degree of multi-tasking [6] and sustained prolonged vigilant monitoring [7]. The negative effects of the technology-intense ICU environment may hinder nurses’ ability to monitor and signal changes in critically ill patients.

Due to the complexity and fragility of the critically ill patient clinicians need to use different technologies to get a sense of organ function, the physiological systems affected and the overall patient status. The use of multiple technologies, used simultaneously to continually assess the patient status, is termed “multimodal monitoring” [8]. Practically, multimodal monitoring is challenging since nurses must constantly scan each discrete monitoring technology to mentally integrate the data, assess current stability and predict the future trend of the patient to anticipate interventions. In the modern technology-driven ICU, a critical care nurse spends half of the time assimilating information embedded in clinical information systems and 15% of the time on monitoring live vitals [9]. Thus, these aforementioned factors make continuous monitoring during extended periods of time challenging and increase the difficulty of making critical decisions based on large data sets. Nurses’ workload could
potentially be decreased by integrating data into one trend monitoring software from which data is easily retrieved and visualized by the nurses through their interaction with the display interface.

Such data integrating and visualization software for continuous multimodal monitoring has been developed and is the subject of this study. Specifically, we evaluated software entitled “T3™”, which stands for “Tracking, Trajectory and Triggering” and which has been implemented in several North American intensive care units. The software combines all compatible data streams from multimodal monitoring and displays, in real-time, the patient’s historical trends over the entire length of stay, (e.g. days, weeks or months) on a highly interactive and responsive user interface. It has been developed to visualize large quantities of continuous multimodal monitoring data and aid in determining patient risk [10,11] but was originally developed to support physician intensivists. The software interface consists of four main screens: login, unit-level patient census, individual patient trend information and frequently-asked questions (FAQ). The general navigation sequence is shown in Figure 1. Although T3™ has the potential to improve the predictability and reliability of nurses’ decision-making, the design of any medical technology’s interface may lead to incorrect decision-making or worse, create new sources of errors [12] by hindering easy information retrieval, appropriate display of data or contributing to overloading memory capacity. To minimize the potential for user error, the usability, efficiency, and effectiveness of the interface should be assessed.

In the present study, we discuss an expedited method that is commonly used to evaluate the usability of user interfaces, called a heuristic evaluation. Specifically, the evaluation assesses whether aspects of a design are in agreement or in violation of established usability (i.e., ease-of-use) principles, or heuristics [13]. Data resulting from this evaluation can then be used to iteratively redesign the interface.

Several sets of heuristics have been proposed in literature, and their application has been extended beyond software interface evaluation. For instance, these heuristics have been modified for and applied to several medical device interfaces [14]. Heuristic evaluations are conducted by people that have expertise in human factors and sometimes with the help of an expert knowledge user. Typically, two or three evaluators independently conduct the evaluation and identify usability issues.

In sum, this present study aimed to demonstrate the use of heuristic evaluation to assess and improve current and future continuous monitoring software for intensive care. Results of this evaluation are applicable to manufacturers and clinicians wishing to improve the user interface through design of these and other healthcare monitoring systems.

Materials and Methods

Setting

The data integration and display software was launched at the pediatric intensive care and cardiac critical care units of a large academic hospital, in Canada. Together these intensive care units, on the same floor, contain 36 beds and are equally distributed between the two units. There are single and multiple patient rooms, and each

Figure 1: The four main screens of the integrating software
The bedspot is equipped with the same patient monitoring system charting system.

**Data integration and visualization software**

In this study, T3™ version 1.6 was evaluated. At the time of the evaluations, the signals which could be visualized were the basic vitals, end-tidal CO2 (integrated in 2013), intracranial pressure, and others listed in Table 1. The display includes these abbreviations and more based on the monitors connected to the patient. Collectively, they represent several discrete locations which include the physiological monitor above the bedside, sometimes the mechanical ventilator and any of three vendor-specific versions of near infrared spectrometers. As of July 2015, near-infrared spectroscopy (NIRS) signals, such as regional oxygen saturation (rSO2), were integrated into the software as part of one of the research group's goals of comprehensively integrating continuous monitoring signals, and reducing signal redundancy.

<table>
<thead>
<tr>
<th>Patient signals</th>
<th>Signal Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart Rate</td>
<td>HR</td>
</tr>
<tr>
<td>Respiratory Rate</td>
<td>Resp</td>
</tr>
<tr>
<td>Pulse</td>
<td>Pulse</td>
</tr>
<tr>
<td>Percent oxygen saturation</td>
<td>SpO2</td>
</tr>
<tr>
<td>Non-Invasive Blood Pressure (systolic, mean or diastolic)</td>
<td>NBPs, NBPm, NBPd</td>
</tr>
<tr>
<td>Arterial Blood Pressure (systolic, mean or diastolic)</td>
<td>ABPs, ABPm, ABPd</td>
</tr>
<tr>
<td>Airway respiratory rate</td>
<td>awRR</td>
</tr>
<tr>
<td>Temperature</td>
<td>T</td>
</tr>
<tr>
<td>Central Venous Pressure</td>
<td>CVP</td>
</tr>
<tr>
<td>Intracranial Pressure</td>
<td>ICP</td>
</tr>
<tr>
<td>End-tidal Carbon Dioxide</td>
<td>etCO2</td>
</tr>
<tr>
<td>Inspired minimum Carbon Dioxide</td>
<td>imCO2</td>
</tr>
<tr>
<td>Regional Oxygen Saturation</td>
<td>rSO2</td>
</tr>
</tbody>
</table>

**Table 1: List of selected patient signals viewable on the data integrating and visualization software**

<table>
<thead>
<tr>
<th>#</th>
<th>Code</th>
<th>Heuristic</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Consistency</td>
<td>Consistency and Standards</td>
<td>Users should not have to wonder whether different words, situations, or actions mean the same thing. Standards and conventions in product design should be followed.</td>
</tr>
<tr>
<td>2</td>
<td>Visibility</td>
<td>Visibility of System State</td>
<td>Visibility of system state. Users should be informed about what is going on with the system through appropriate feedback and display of information.</td>
</tr>
<tr>
<td>3</td>
<td>Match</td>
<td>Match Between System and World</td>
<td>The image of the system perceived by users should match the model the users have about the system.</td>
</tr>
<tr>
<td>4</td>
<td>Minimalist</td>
<td>Minimalist</td>
<td>Any extraneous information is a distraction and a slow-down.</td>
</tr>
<tr>
<td>5</td>
<td>Memory</td>
<td>Minimize Memory Load</td>
<td>Minimize memory load. Users should not be required to memorize a lot of information to carry out tasks. Memory load reduces users capacity to carry out the main tasks.</td>
</tr>
</tbody>
</table>

Nurses can view both patients in the current census (ICU patient population) and previously discharged patients in the archive database. The patient screen is where all continuous monitored signals, as well as intermittent signals, such as non-invasive blood pressures, can be viewed on a single screen.

**Heuristic Evaluation: Applying Usability Heuristics for Medical Devices**

The heuristic evaluation was conducted in three rounds: one in December 2013 and two in May 2014. During these evaluation rounds, three evaluators assessed the same version of the software for usability issues. In the first round, one "double-specialist" with novice-level knowledge of both the clinical work and human factors assessed the interface. In the second round, one domain expert from bedside clinical nursing and another domain expert from human factors together assessed the interface. A short third round to evaluate the interface in the clinical setting was performed by the single "double-specialist" of the first round.

In the two first rounds, the software was viewed on a 15” Samsung Series 9 laptop, with screen resolution of 1600 x 9000, 8GB of memory and an Intel Core i7-3517U central processing unit, running Windows 8 64-bit operating system, connected to the internal network and accessing the day’s patient census and their continuously monitored signals.

The interface was assessed using 14 heuristics, or “rules of thumb”, developed by leading experts in interface design and modified for medical devices [14-16], see Table 2 for the complete list. When conducting a heuristic evaluation, each usability issue is described, along with which heuristic(s) it violates and the potential impact it can have. Usability issues often are associated with more than one type of heuristic violation; these issues are then rated for severity (0: cosmetic to 4: usability catastrophe, see Table 3). The results of the two rounds were pooled; in case of discrepancy they were discussed between the human factors researchers who each participated in the evaluation rounds and consensus on heuristic violations and severity was reached.

The potential clinical impact of the issues, in the clinical setting, was confirmed with a medical domain expert and frequent user of the software.
Feedback

Informative Feedback

Users should be given prompt and informative feedback about their actions.

Flexibility

Flexibility and Efficiency

Users always learn and users are always different. Give users the flexibility of creating customization and shortcuts to accelerate their performance.

Message

Good Error Message

The messages should be informative enough such that users can understand the nature of errors, learn from errors, and recover from errors.

Error

Prevent Error

It is always better to design interfaces that prevent errors from happening in the first place.

Closure

Clear Closure

Every task has a beginning and an end. Users should be clearly notified about the completion of a task.

Undo

Reversible actions

Users should be allowed to recover from errors. Reversible actions also encourage exploratory learning.

Language

Use Users’ Language

The language should be always presented in a form understandable by the intended users.

Control

Users in Control

Do not give users that impression that they are controlled by the systems.

Document

Help and documentation

Always provide help when needed.

Table 2: The 14 usability heuristics for medical devices as defined by Zhang et al.[14]

<table>
<thead>
<tr>
<th>Severity</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>not a usability problem at all</td>
</tr>
<tr>
<td>1</td>
<td>cosmetic problem only</td>
</tr>
<tr>
<td>2</td>
<td>minor usability problem</td>
</tr>
<tr>
<td>3</td>
<td>major usability problem</td>
</tr>
<tr>
<td>4</td>
<td>usability catastrophe</td>
</tr>
</tbody>
</table>

Table 3: Severity rating as defined by Zhang et al.[14]

Results

In total, 50 usability issues were found. Two percent of usability issues were rated as a catastrophic problem (severity = 4), 38% were rated as major usability problems (severity = 3), 56% were rated as minor usability problems (severity = 2), and 4% were cosmetic usability problems (severity = 1).

The 50 usability issues were associated with 194 heuristic violations, as shown in Figure 2. The most common types of heuristic violations, with over 15 occurrences, were memory, visibility, match, error, minimalist, and flexibility. The “double-expert” team, consisting of a senior critical care nurse and a human factors expert, revealed 49 more violations than the single “double-specialist” evaluator and attributed severity to more heuristic violations. When severity for all issues, from both rounds, was compared there was a 68% severity rating match between the two.

Most important issues which should be addressed were the manipulation of the timeline (severity of 4 - usability catastrophe), use of shading to highlight signals which were out of range, and lack of an undo function (severity of 3 - major usability problem). These examples and others are discussed below.

Example #1 - Catastrophic problem

**Issue description:** The most important usability issue involved choosing the timeframe of data to be viewed and was rated as a catastrophic problem. Selecting the timeframe of data to be viewed is an important task as nurses are often required to compare a patient's stability (i.e., vitals are within a target range) at a given point in time to the patient's baseline values observed at another point in time. For example, ICU nurses who often temporarily care for other nurses' patients, when covering during breaks, may choose to review the
patient vitals from the previous hours to get a sense of the patient’s stability over time. To do so, the nurse would need to interact with the software interface and specify the time frame of continuous patient data s/he would like to view. However, s/he may encounter difficulty when trying to choose the timeframe because the icons are very small requiring high visual acuity and dexterity with the mouse cursor to select the desired timeframe. Not being able to easily manipulate the timeframe could lead to faulty decision making since interpreting the patient data requires correct time orientation (e.g. start and end of data, time period of data, relative time period). Thus, the usability of timescale manipulation is critical since its potential impact on clinical practice is high. In Figure 3, the illustrative example shows one way to choose the time period of the data.

Heuristics violated: Consistency; visibility; match; memory; minimalist; memory; feedback; error; undo and control.

Recommendation: Users should be able to manipulate the trend data in a way that they feel in control of their selection and can easily identify what they have selected. The timeframe of the data window should be more apparent, say with a larger sized font.

Example #2 - Major usability issue

Issue description: A major usability issue was the use of shading as an aid to rapidly visualize patient signals that are out of range. Rapid visualization of out-of-range patient signals is a critical feature because it can indicate duration and severity of patient stability. Although shading of a single parameter may be clearly seen and understood, this feature may lead to confusion when several patient signal trends are viewed on the same graph. Specifically when multiple signals are viewed, the various shadings may overlap thereby hindering nurses' ability to detect which specific signal or signals should be addressed. Such confusion could lead to inappropriate interventions potentially causing patient harm. Figure 4 shows overlapped multiple signals, each with different coloured shadings.
Heuristics Violated: Visibility; memory; feedback and error.

Recommendation: Users should be able to interpret patient instability and detect which specific signal is unstable, without having to rely on their memory to understand visualization cues such as shading. When it is desirable to view many patient signals and their targets ranges on the same graph, consider cues other than shading to rapidly identify which signals are out-of-range.

Example #3 - Major usability issue

Issue description: No “undo” for many actions including zooming in (i.e., no zooming out), moving the time window along the timescale, and dragging-and-dropping several variables on one graph. The absence of this function discourages exploration and learning, and could lead to error in time sensitive situations.

Heuristics Violated: Consistency; match; memory; flexibility; error; undo and control.

Recommendation: New users should be able to perform actions and reversible actions to learn through exploration. More importantly, when manipulating the interface to visualize data, if an action creates a worse representation, users should be able to go back to a previous configuration rather than start from a default setting or an inappropriate configuration. Frequent users should be able to reverse actions to prevent serious errors or unintentional data representation. Designers should consider programming an “undo” command for several of the functionalities mentioned in this issue’s description and as a standard command for any actions performed at the interface.

Example #4 - Minor usability issue

Issue description: Use of words that hold different or no meaning to nurses in their clinical practice. For example, in the census, the column “First Message” appears but does not relate to information useful to their clinical decision-making. Also in the census, discharged patient data are located in the “Archived patients” census. Another example is the use of computer programming terms such as “Administrator” and “Modifier”, in the FAQ, which are specialized terms for computer programmers but may not be understood by clinicians.

Heuristics Violated: Match; memory and language

Recommendations: Change or eliminate the words or information which are unfamiliar to clinicians.

Example #5 - Positive Features

In this sixth iteration, the software interface uses design elements that have been recognized as helpful to end-users. First, the right-hand legend provides the choice of 5 minutes, 30 minutes or 12-hour trends and are similar to sparklines, developed by Tuftes, and are described as “data-intense, design-simple, word-sized graphics” [17]. In a clinical setting, these “sparklines” (i.e., small representation of data) were found to be useful in providing physicians with trend information [18].

Example #6 - Positive Features

Second, a design feature that adhered well to the heuristics of consistency and match was the default colours for traces of heart rate,
blood pressures and oxygen saturation. Specifically, the colors chosen to represent these vitals, on the T³™ interface, matched the colours used by the bedside physiological monitor. Although no standard exists to represent physiological variables, the colors used by the T³™ software matched those used in this study’s ICU setting and nurses were familiar with them. In practice, when switching from the T³™ display back to the physiological monitor, identifying the traces based on color would require minimal cognitive effort due to adherence to match and consistency heuristics.

**Discussion**

From the heuristic evaluation, 40% of the usability issues identified were categorized as major or catastrophic usability issues and the remainder, that is 60%, were minor or cosmetic usability problems. Collectively, the major and catastrophic usability issues could have serious impact on patient safety and should be addressed. In particular, timescale manipulation was identified as a catastrophic issue with physiological data representation. Past research has shown that such timescale manipulation issues contributed to physicians’ and nurses’ inability to see when a particular physiological parameter has reached a critical point [19]. Therefore the catastrophic problem of time manipulation requires much attention given the round the clock nature of critical care.

The three most violated heuristics were those of “memory”, “visibility” and “match”. This indicates the need to 1) design the software so that using it minimizes cognitive load, 2) display information which clearly indicates what the system is doing, and 3) ensure the interface displays trend information using cues familiar to nurses.

As the T³™ system integrates more of the monitoring technologies (e.g., electroencephalogram) and even therapeutic technologies (e.g., infusion pumps, ventilators and feeding pumps) its impact on decision making will extend to many other clinicians (e.g. pharmacists, respiratory therapists and dieticians) who may have different interface requirements. The software’s extended use to the different types of clinicians could eventually lead to an impact on team-based clinical decision-making. Thus, consideration must be given to the expected usability issues due to medical device integration and use with other clinical information systems. That is, continual efforts to integrate more of the stand-alone medical devices into this display may create new usability issues as more patient signals are visualized. Designers should consider the heuristics for medical devices, in the context of the changing multimodal monitoring system and advances in clinical instrumentation. In addition, as new signals, features and functions are added to the software, these may impact the interface layout and adherence to the core heuristics. For example, a possible usability issue may be the visualization of intermittent non-invasive blood pressures in addition to the continuous invasive blood pressure. The ability to visualize a new type of blood pressure, in the form of non-continuous data points, may pose a visualization challenge. To avoid confusion, a quick heuristic evaluation when a new type of data is integrated into the software is recommended.

Another issue is the level of detail of the trend information available at the bedside; in this case, a higher level of detail is available from the bedside physiological monitor. The T³™ display aims to provide long-term trend information (e.g. minutes, hours or days, with a minimum of 0.2 second intervals). This information requirement may indicate that any new trend monitoring software must provide progressive level of detail to the waveform-level or make this information available. The choice may not be for one or the other but to have both trends on the same screen or near each other for quick patient baseline comparison. This usability issue may be confirmed through usability testing or simulated clinical decision-making experiments with nurses.

This study represents the first heuristic evaluation of clinically available, highly interactive, data integration and visualization software. The usability issues found through the heuristic evaluation required little cost and the time of one representative end-user (expert nurse) and two human factors researchers, one of which had observed the ICU and staff for eight months prior to the first assessment. When all issues were pooled there was a 68% match of severity ratings. In all instances, severity ratings deviated by one level suggesting use of a three-point (e.g., high-, medium- or low-) severity scale rather than the four-point (i.e., 1-, 2-, 3-, 4-) severity scale may minimize disagreement. Given the potential high-risk, high-impact nature of critical care, the three-point scale would indicate that high and medium severity issues should be addressed and little gain is achieved with categorizing into one more severity level.

Fifty usability issues were found and two positive design features were highlighted. When addressing the usability issues efforts should be made to retain the positive design features. These issues have been shared with the software developers and already some of these issues have been addressed. In the future, we recommend that heuristic evaluations be performed on the user interface before software implementation in the clinical setting.

**Limitations**

This study was highly institutional context-dependant and user-dependant. Three evaluators, divided into one “double-expert” team consisting of one domain expert from nursing and one domain expert from human factors, and one “double-expert”, with intermediate knowledge of both domains may satisfy Nielsen’s requirement of at least two to three double specialists to uncover between 81 to 90% of usability problems [13]. This criterion may not hold for software interfaces used in complex settings and used by several types of users.

Further study should include the involvement of nurses as they use the software to perform tasks, confirming these usability issues and observing many other usability issues occurring with actual use. A subsequent phase involving user physicians, nurses and respiratory therapists is planned.

The heuristic evaluation is meant to be a first step in the iterative user-centred design process. Its strength as a quick evaluation tool means it can be applied as a change-driven process for quick prototyping in view of optimizing the interface before testing with actual users and different types of critical care specialists.

**Conclusion**

The heuristic evaluation method applied by the complimentary team identified and prioritized key interface problems according to severity and impact of the usability issues which can be addressed during the iterative design life cycle of the software. Heuristic violations help guide designers by specifying what type of solution is required and help match solutions with known visualization aids. By
using the decades of knowledge from software interface design and the heuristics for medical devices, basic usability issues were quickly identified with time of few evaluators. Multidisciplinary teams consisting of actual end-users reveal many more usability issues than with single evaluators. Throughout the development of the data integrating and visualization software, quickly finding and addressing the interface usability issues early can facilitate the transition and integration of these systems into the actual setting. This new software tool has the potential to minimize the sources of disparate data and help critical care nurses manage the numerous patient data signals, but the many usability issues must be addressed to minimize potential use errors and realize its full potential.

Acknowledgements

The authors would like to recognize the invaluable participation of the pediatric intensive care nurse who provided the nurse's perspective. We would also like to thank the research teams HumanEra at the Centre for Global eHealth Innovation, University Health Network, and the Guerguerian lab at the Hospital for Sick Children.

The authors would also like to thank our sponsors: the Natural Sciences and Engineering Research Council (NSERC) of Canada, CREATE Academic Rehabilitation Engineering (CARE) training program, Barbara and Frank Milligan's support of biomedical engineering research through their graduate Fellowship, and the Canadian Institutes of Health Research's Healthcare, Technology and Place (HCTP) strategic initiative.

References